(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/46407 A2

(51) International Patent Classification[7]: C12N 15/12, 15/11, 15/63, 1/21, 5/10, C07K 14/705, 16/28, A01K 67/027, C12Q 1/68, G01N 33/68

(21) International Application Number: PCT/US01/45661

(22) International Filing Date: 5 December 2001 (05.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/729,094 5 December 2000 (05.12.2000) US

(71) Applicant: **PE CORPORATION (NY)** [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: **CHATURVEDI, Kabir**; c/o Celera Genomics Corporation, 45 West Gude Drive C2-4 #21, Rockville, MD 20850 (US). **WEI, Ming-Hui**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **KETCHUM, Karen, A.**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **DIFRANCESCO, Valentina**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **BEASLEY, Ellen, M.**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF

```
   1 GTCTCCCTCC CGCGCGATGG CCTCGGCGCT GAGCTATGTC TCCAAGTTCA
  51 AGTCCTTCGT GATCTTGTTC GTCACCCCGC TCCTGCTGCT GCCACTCGTC
 101 ATTCTGATGC CCGCCAAGTT TGTCAGGTGT GCCTACGTCA TCATCCTCAT
 151 GGCCATTTAC TGGTGCACAG AAGTCATCCC TCTGGCTGTC ACCTCTCTCA
 201 TGCCTGTCTT GCTTTTCCCA CTCTTCCAGA TTCTGGACTC CAGGCAGGTG
 251 TGTGTCCAGT ACATGAAGGA CACCAACATG CTGTTCCTGG GCGGCCTCAT
 301 CGTGGCCGTG GCTGTGGAGC GCTGGAACCT GCACAAGAGG ATCGCCCTGC
 351 GCACGCTCCT CTGGGTGGGG GCCAAGCCTG CACGGCTGAT GCTGGGCTTC
 401 ATGGGCGTCA CAGCCCTCCT GTCCATGTGG ATCAGTAACA TGGCAACCAC
 451 GGCCATGATG GTGCCCATCG TGGAGGCCAT ATTGCAGCAG ATGGAAGCCA
 501 CAAGCGCAGC CACCGAGGCC GGCCTGGAGC TGGTGGACAA GGGCAAGGCC
 551 AAGGAGCTGC CAGGGAGTCA AGTGATTTTT GAAGGCCCCA CTCTGGGGCA
 601 GCAGGAAGAC CAAGAGCGGA AGAGGTTGTG TAAGGCCATG ACCCTGTGCA
 651 TCTGCTACGC GGCCAGCATC GGGGGCACCG CCACCCTGAC CGGGACGGGA
 701 CCCAACGTGG TGCTCCTGGG CCAGATGAAC GAGTTGTTTC CTGACAGCAA
 751 GGACCTCGTG AACTTTGCTT CCTGGTTTGC ATTTGCCTTT CCCAACATGC
 801 TGGTGATGCT GCTGTTCGCC TGGCTGTGGC TCCAGTTTGT TTACATGAGA
 851 TTCAATTTTA AAAAGTCCTG GGGCTGCGGG CTAGAGAGCA AGAAAAACGA
 901 GAAGGCTGCC CTCAAGGTGC TGCAGGAGGA GTACCGGAAG CTGGGGCCCT
 951 TGTCCTTCGC GGAGATCAAC GTGCTGATCT GCTTCTTCCT GCTGGTCATC
1001 CTGTGGTTCT CCCGAGACCC CGGCTTCATG CCCGGCTGGC TGACTGTTGC
1051 CTGGGTGGAG GGTGAGACAA AGTATGTCTC CGATGCCACT GTGGCCATCT
1101 TTGTGGCCAC CCTGCTATTC ATTGTGCTTT CACAGAAGCC CAAGTTTAAC
1151 TTCCGCAGCC AGACTGAGGA AGAAAGGAAA ACTCCATTTT ATCCCCCTCC
1201 CCTGCTGGAT TGGAAGGTAA CCCAGGAGAA AGTGCCCTGG GGCATCGTGC
1251 TGCTACTAGG GGGCGGATTT GCTCTGGCTA AAGGATCCGA GGCCTCGGGG
1301 CTGTCCGTGT GGATGGGGAA GCAGATGGAG CCCTTGCACG CAGTGCCCCC
1351 GGCAGCCATC ACCTTGATCT TGTCCTTGCT CGTTGCCGTG TTCACTGAGT
1401 GCACAAGCAA CGTGGCCACC ACCACCTTGT TCCTGCCCAT CTTTGCCTCC
1451 ATGTCTCGCT CCATCGGCCT CAATCCGCTG TACATCATGC TGCCCTGTAC
1501 CCTGAGTGCC TCCTTTGCCT TCATGTTGCC TGTGGCCACC CCTCCAAATG
1551 CCATCCTGTT CACCTATGGG CACCTCAAGG TTGCTGACAT GGTGAAAACA
1601 GGAGTCATAA TGAACATAAT TGGAGTCTTC TGTGTGTTTT TGGCTGTCAA
1651 CACCTGGGGA CGGGCCATAT TTGACTTGGA TCATTTCCCT GACTGGGCTA
1701 ATGTGACACA TATTGAGACT TAGGAAGAGC CACAAGACCA CACACACAGC
1751 CCTTACCCTC CTCAGGACTA CCGAACCTTC TGGCACACCT TGTACAGAGT
1801 TTTGGGGTTC ACACCCCAAA ATGACCCAAC GATGTCCACA CACCACCAAA
1851 ACCCAGCCAA TGGGCCACCT CTTCCTCCAA GCCCAGATGC AGAGATGGTC
1901 ATGGGCAGCT GGAGGGTAGG CTCAGAAATG AAGGGAACCC CTCAGTGGGC
1951 TGCTGGACCC ATCTTTCCCA AGCCTTGCCA TTATCTCTGT GAGGGAGGCC
2001 AGGTAGCCGA GGGATCAGGA TGCAGGCTGC TGTACCCGCT CTGCCTCAAG
2051 CATCCCCCAC ACAGGGCTCT GGTTTTCACT CGCTTCGTCC TAGATAGTTT
2101 AAATGGGAAT CAGATCCCCT GGTTGAGAGC TAAGACAACC ACCTACCAGT
2151 GCCCATGTCC CTTCCAGCTC ACCTTGAGCA GCCTCAGATC ATCTCTGTCA
2201 CTCTGGAAGG GACACCCCAG CCA (SEQ ID NO:1)

FEATURES:
5'UTR:       1 - 16
Start Codon: 17
Stop Codon:  1721
3'UTR:       1724
```

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the transporter peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the transporter peptides, and methods of identifying modulators of the transporter peptides.

WO 02/46407 A2

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF

RELATED APPLICATIONS

The present application claims priority to U.S. Application No. 09/729,094, filed December 5, 2000 (Atty. Docket CL000662).

FIELD OF THE INVENTION

The present invention is in the field of transporter proteins that are related to the sodium-dependent dicarboxylate transporter subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel peptides and proteins that effect ligand transport and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

Transporters

Transporter proteins regulate many different functions of a cell, including cell proliferation, differentiation, and signaling processes, by regulating the flow of molecules such as ions and macromolecules, into and out of cells. Transporters are found in the plasma membranes of virtually every cell in eukaryotic organisms. Transporters mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of molecules and ion across cell membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, transporters, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) Annu. Rev. Physiol. 50:111-122.

Transporters are generally classified by structure and the type of mode of action. In addition, transporters are sometimes classified by the molecule type that is transported, for example, sugar transporters, chlorine channels, potassium channels, etc. There may be many classes of channels for transporting a single type of molecule (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters: Receptor and transporter nomenclature supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 (1997) and http://www-biology.ucsd.edu/~msaier/transport/titlepage2.html.

The following general classification scheme is known in the art and is followed in the present discoveries.

Channel-type transporters. Transmembrane channel proteins of this class are ubiquitously found in the membranes of all types of organisms from bacteria to higher eukaryotes. Transport systems of this type catalyze facilitated diffusion (by an energy-independent process) by passage through a transmembrane aqueous pore or channel without evidence for a carrier-mediated mechanism. These channel proteins usually consist largely of a-helical spanners, although b-strands may also be present and may even comprise the channel. However, outer membrane porin-type channel proteins are excluded from this class and are instead included in class 9.

Carrier-type transporters. Transport systems are included in this class if they utilize a carrier-mediated process to catalyze uniport (a single species is transported by facilitated diffusion), antiport (two or more species are transported in opposite directions in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy) and/or symport (two or more species are transported together in the same direction in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy).

Carrier-type transporters include the Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter ("DAACS") family, which catalyze Na+ and/or H+ symport together with (a) a Krebs cycle dicarboxylate (malate, succinate, or fumarate), (b) a dicarboxylic amino acid (glutamate or aspartate), (c) a small, semipolar, neutral amino acid (Ala, Ser, Cys, Thr), (d) both neutral and acidic amino acids or (e) most zwitterionic and dibasic amino acids. The bacterial members are of about 450 (420-491) amino acyl residues while the mammalian proteins are of about 550 (503-574) residues in length. These proteins possess between ten and twelve putative transmembrane spanners (TMSs). A specific topological model in which 7 a-helical TMSs are followed by a reentrant loop-pore structure followed by one final TMS is presented in Slotboom et al., *Microbiol. Mol. Biol. Rev.* 63: 293-3071999 (1999). All of the bacterial proteins cluster together on the phylogenetic tree as do the mammalian proteins. The mammalian permeases that transport neutral amino acids cluster separately from those that are specific for the acidic amino acids. Among the mammalian proteins are neuronal excitatory amino acid neurotransmitter permeases.

Pyrophosphate bond hydrolysis-driven active transporters. Transport systems are included in this class if they hydrolyze pyrophosphate or the terminal pyrophosphate bond in ATP or another nucleoside triphosphate to drive the active uptake and/or extrusion of a solute or solutes. The transport protein may or may not be transiently phosphorylated, but the substrate is not phosphorylated.

PEP-dependent, phosphoryl transfer-driven group translocators. Transport systems of the bacterial phosphoenolpyruvate:sugar phosphotransferase system are included in this class. The product of the reaction, derived from extracellular sugar, is a cytoplasmic sugar-phosphate.

Decarboxylation-driven active transporters. Transport systems that drive solute (e.g., ion) uptake or extrusion by decarboxylation of a cytoplasmic substrate are included in this class.

Oxidoreduction-driven active transporters. Transport systems that drive transport of a solute (e.g., an ion) energized by the flow of electrons from a reduced substrate to an oxidized substrate are included in this class.

Light-driven active transporters. Transport systems that utilize light energy to drive transport of a solute (e.g., an ion) are included in this class.

Mechanically-driven active transporters. Transport systems are included in this class if they drive movement of a cell or organelle by allowing the flow of ions (or other solutes) through the membrane down their electrochemical gradients.

Outer-membrane porins (of b-structure). These proteins form transmembrane pores or channels that usually allow the energy independent passage of solutes across a membrane. The transmembrane portions of these proteins consist exclusively of b-strands that form a b-barrel. These porin-type proteins are found in the outer membranes of Gram-negative bacteria, mitochondria and eukaryotic plastids.

Methyltransferase-driven active transporters. A single characterized protein currently falls into this category, the Na+-transporting methyltetrahydromethanopterin:coenzyme M methyltransferase.

Non-ribosome-synthesized channel-forming peptides or peptide-like molecules. These molecules, usually chains of L- and D-amino acids as well as other small molecular building blocks such as lactate, form oligomeric transmembrane ion channels. Voltage may induce channel formation by promoting assembly of the transmembrane channel. These peptides are often made by bacteria and fungi as agents of biological warfare.

Non-Proteinaceous Transport Complexes. Ion conducting substances in biological membranes that do not consist of or are not derived from proteins or peptides fall into this category.

Functionally characterized transporters for which sequence data are lacking. Transporters of particular physiological significance will be included in this category even though a family assignment cannot be made.

Putative transporters in which no family member is an established transporter. Putative transport protein families are grouped under this number and will either be classified elsewhere

when the transport function of a member becomes established, or will be eliminated from the TC classification system if the proposed transport function is disproven. These families include a member or members for which a transport function has been suggested, but evidence for such a function is not yet compelling.

Auxiliary transport proteins. Proteins that in some way facilitate transport across one or more biological membranes but do not themselves participate directly in transport are included in this class. These proteins always function in conjunction with one or more transport proteins. They may provide a function connected with energy coupling to transport, play a structural role in complex formation or serve a regulatory function.

Transporters of unknown classification. Transport protein families of unknown classification are grouped under this number and will be classified elsewhere when the transport process and energy coupling mechanism are characterized. These families include at least one member for which a transport function has been established, but either the mode of transport or the energy coupling mechanism is not known.

Ion channels

An important type of transporter is the ion channel. Ion channels regulate many different cell proliferation, differentiation, and signaling processes by regulating the flow of ions into and out of cells. Ion channels are found in the plasma membranes of virtually every cell in eukaryotic organisms. Ion channels mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of ion across epithelial membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, ion channels, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) Annu. Rev. Physiol. 50:111-122.

Ion channels are generally classified by structure and the type of mode of action. For example, extracellular ligand gated channels (ELGs) are comprised of five polypeptide subunits, with each subunit having 4 membrane spanning domains, and are activated by the binding of an extracellular ligand to the channel. In addition, channels are sometimes classified by the ion type that is transported, for example, chlorine channels, potassium channels, etc. There may be many classes of channels for transporting a single type of ion (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters (1997). Receptor and ion channel nomenclature supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 and http://www-biology.ucsd.edu/~msaier/transport/toc.html.

There are many types of ion channels based on structure. For example, many ion channels fall within one of the following groups: extracellular ligand-gated channels (ELG), intracellular ligand-gated channels (ILG), inward rectifying channels (INR), intercellular (gap junction) channels, and voltage gated channels (VIC). There are additionally recognized other channel families based on ion-type transported, cellular location and drug sensitivity. Detailed information on each of these, their activity, ligand type, ion type, disease association, drugability, and other information pertinent to the present invention, is well known in the art.

Extracellular ligand-gated channels, ELGs, are generally comprised of five polypeptide subunits, Unwin, N. (1993), Cell 72: 31-41; Unwin, N. (1995), Nature 373: 37-43; Hucho, F., et al., (1996) J. Neurochem. 66: 1781-1792; Hucho, F., et al., (1996) Eur. J. Biochem. 239: 539-557; Alexander, S.P.H. and J.A. Peters (1997), Trends Pharmacol. Sci., Elsevier, pp. 4-6; 36-40; 42-44; and Xue, H. (1998) J. Mol. Evol. 47: 323-333. Each subunit has 4 membrane spanning regions: this serves as a means of identifying other members of the ELG family of proteins. ELG bind a ligand and in response modulate the flow of ions. Examples of ELG include most members of the neurotransmitter-receptor family of proteins, e.g., GABAI receptors. Other members of this family of ion channels include glycine receptors, ryandyne receptors, and ligand gated calcium channels.

The Voltage-gated Ion Channel (VIC) Superfamily

Proteins of the VIC family are ion-selective channel proteins found in a wide range of bacteria, archaea and eukaryotes Hille, B. (1992), Chapter 9: Structure of channel proteins; Chapter 20: Evolution and diversity. In: Ionic Channels of Excitable Membranes, 2nd Ed., Sinaur Assoc. Inc., Pubs., Sunderland, Massachusetts; Sigworth, F.J. (1993), Quart. Rev. Biophys. 27: 1-40; Salkoff, L. and T. Jegla (1995), Neuron 15: 489-492; Alexander, S.P.H. et al., (1997), Trends Pharmacol. Sci., Elsevier, pp. 76-84; Jan, L.Y. et al., (1997), Annu. Rev. Neurosci. 20: 91-123; Doyle, D.A, et al., (1998) Science 280: 69-77; Terlau, H. and W. Stühmer (1998), Naturwissenschaften 85: 437-444. They are often homo- or heterooligomeric structures with several dissimilar subunits (e.g., a1-a2-d-b Ca^{2+} channels, ab_1b_2 Na^+ channels or $(a)_4$-b K^+ channels), but the channel and the primary receptor is usually associated with the a (or a1) subunit. Functionally characterized members are specific for K^+, Na^+ or Ca^{2+}. The K^+ channels usually consist of homotetrameric structures with each a-subunit possessing six transmembrane spanners (TMSs). The a1 and a subunits of the Ca^{2+} and Na^+ channels, respectively, are about four times as large and possess 4 units, each with 6 TMSs separated by a hydrophilic loop, for a total of 24 TMSs. These large channel proteins form heterotetra-unit structures equivalent to the homotetrameric structures of most K^+ channels. All four units of the Ca^{2+} and Na^+ channels are

homologous to the single unit in the homotetrameric K^+ channels. Ion flux via the eukaryotic channels is generally controlled by the transmembrane electrical potential (hence the designation, voltage-sensitive) although some are controlled by ligand or receptor binding.

Several putative K^+-selective channel proteins of the VIC family have been identified in prokaryotes. The structure of one of them, the KcsA K^+ channel of *Streptomyces lividans*, has been solved to 3.2 Å resolution. The protein possesses four identical subunits, each with two transmembrane helices, arranged in the shape of an inverted teepee or cone. The cone cradles the "selectivity filter" P domain in its outer end. The narrow selectivity filter is only 12 Å long, whereas the remainder of the channel is wider and lined with hydrophobic residues. A large water-filled cavity and helix dipoles stabilize K^+ in the pore. The selectivity filter has two bound K^+ ions about 7.5 Å apart from each other. Ion conduction is proposed to result from a balance of electrostatic attractive and repulsive forces.

In eukaryotes, each VIC family channel type has several subtypes based on pharmacological and electrophysiological data. Thus, there are five types of Ca^{2+} channels (L, N, P, Q and T). There are at least ten types of K^+ channels, each responding in different ways to different stimuli: voltage-sensitive [Ka, Kv, Kvr, Kvs and Ksr], Ca^{2+}-sensitive [BK_{Ca}, IK_{Ca} and SK_{Ca}] and receptor-coupled [K_M and K_{ACh}]. There are at least six types of Na^+ channels (I, II, III, μ1, H1 and PN3). Tetrameric channels from both prokaryotic and eukaryotic organisms are known in which each a-subunit possesses 2 TMSs rather than 6, and these two TMSs are homologous to TMSs 5 and 6 of the six TMS unit found in the voltage-sensitive channel proteins. KcsA of *S. lividans* is an example of such a 2 TMS channel protein. These channels may include the K_{Na} (Na^+-activated) and K_{Vol} (cell volume-sensitive) K^+ channels, as well as distantly related channels such as the Tok1 K^+ channel of yeast, the TWIK-1 inward rectifier K^+ channel of the mouse and the TREK-1 K^+ channel of the mouse. Because of insufficient sequence similarity with proteins of the VIC family, inward rectifier K^+ IRK channels (ATP-regulated; G-protein-activated) which possess a P domain and two flanking TMSs are placed in a distinct family. However, substantial sequence similarity in the P region suggests that they are homologous. The b, g and d subunits of VIC family members, when present, frequently play regulatory roles in channel activation/deactivation.

The Epithelial Na^+ Channel (ENaC) Family

The ENaC family consists of over twenty-four sequenced proteins (Canessa, C.M., et al., (1994), Nature 367: 463-467, Le, T. and M.H. Saier, Jr. (1996), Mol. Membr. Biol. 13: 149-157; Garty, H. and L.G. Palmer (1997), Physiol. Rev. 77: 359-396; Waldmann, R., et al., (1997), Nature 386: 173-177; Darboux, I., et al., (1998), J. Biol. Chem. 273: 9424-9429; Firsov, D., et

al., (1998), EMBO J. 17: 344-352; Horisberger, J.-D. (1998). Curr. Opin. Struc. Biol. 10: 443-449). All are from animals with no recognizable homologues in other eukaryotes or bacteria. The vertebrate ENaC proteins from epithelial cells cluster tightly together on the phylogenetic tree: voltage-insensitive ENaC homologues are also found in the brain. Eleven sequenced *C. elegans* proteins, including the degenerins, are distantly related to the vertebrate proteins as well as to each other. At least some of these proteins form part of a mechano-transducing complex for touch sensitivity. The homologous *Helix aspersa* (FMRF-amide)-activated Na^+ channel is the first peptide neurotransmitter-gated ionotropic receptor to be sequenced.

Protein members of this family all exhibit the same apparent topology, each with N- and C-termini on the inside of the cell, two amphipathic transmembrane spanning segments, and a large extracellular loop. The extracellular domains contain numerous highly conserved cysteine residues. They are proposed to serve a receptor function.

Mammalian ENaC is important for the maintenance of Na^+ balance and the regulation of blood pressure. Three homologous ENaC subunits, alpha, beta, and gamma, have been shown to assemble to form the highly Na^+-selective channel. The stoichiometry of the three subunits is $alpha_2$, beta1, gamma1 in a heterotetrameric architecture.

The Glutamate-gated Ion Channel (GIC) Family of Neurotransmitter Receptors

Members of the GIC family are heteropentameric complexes in which each of the 5 subunits is of 800-1000 amino acyl residues in length (Nakanishi, N., et al, (1990), Neuron 5: 569-581; Unwin, N. (1993), Cell 72: 31-41; Alexander, S.P.H. and J.A. Peters (1997) Trends Pharmacol. Sci., Elsevier, pp. 36-40). These subunits may span the membrane three or five times as putative a-helices with the N-termini (the glutamate-binding domains) localized extracellularly and the C-termini localized cytoplasmically. They may be distantly related to the ligand-gated ion channels, and if so, they may possess substantial b-structure in their transmembrane regions. However, homology between these two families cannot be established on the basis of sequence comparisons alone. The subunits fall into six subfamilies: a, b, g, d, e and z.

The GIC channels are divided into three types: (1) a-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA)-, (2) kainate- and (3) N-methyl-D-aspartate (NMDA)-selective glutamate receptors. Subunits of the AMPA and kainate classes exhibit 35-40% identity with each other while subunits of the NMDA receptors exhibit 22-24% identity with the former subunits. They possess large N-terminal, extracellular glutamate-binding domains that are homologous to the periplasmic glutamine and glutamate receptors of ABC-type uptake permeases of Gram-negative bacteria. All known members of the GIC family are from animals.

The different channel (receptor) types exhibit distinct ion selectivities and conductance properties. The NMDA-selective large conductance channels are highly permeable to monovalent cations and Ca^{2+}. The AMPA- and kainate-selective ion channels are permeable primarily to monovalent cations with only low permeability to Ca^{2+}.

The Chloride Channel (ClC) Family

The ClC family is a large family consisting of dozens of sequenced proteins derived from Gram-negative and Gram-positive bacteria, cyanobacteria, archaea, yeast, plants and animals (Steinmeyer, K., et al., (1991), Nature 354: 301-304; Uchida, S., et al., (1993), J. Biol. Chem. 268: 3821-3824; Huang, M.-E., et al., (1994), J. Mol. Biol. 242: 595-598; Kawasaki, M., et al, (1994), Neuron 12: 597-604; Fisher, W.E., et al., (1995), Genomics. 29:598-606; and Foskett, J.K. (1998), Annu. Rev. Physiol. 60: 689-717). These proteins are essentially ubiquitous, although they are not encoded within genomes of *Haemophilus influenzae*, *Mycoplasma genitalium,* and *Mycoplasma pneumoniae*. Sequenced proteins vary in size from 395 amino acyl residues *(M. jannaschii)* to 988 residues (man). Several organisms contain multiple ClC family paralogues. For example, *Synechocystis* has two paralogues, one of 451 residues in length and the other of 899 residues. *Arabidopsis thaliana* has at least four sequenced paralogues, (775-792 residues), humans also have at least five paralogues (820-988 residues), and *C. elegans* also has at least five (810-950 residues). There are nine known members in mammals, and mutations in three of the corresponding genes cause human diseases. *E. coli*, *Methanococcus jannaschii* and *Saccharomyces cerevisiae* only have one ClC family member each. With the exception of the larger *Synechocystis* paralogue, all bacterial proteins are small (395-492 residues) while all eukaryotic proteins are larger (687-988 residues). These proteins exhibit 10-12 putative transmembrane a-helical spanners (TMSs) and appear to be present in the membrane as homodimers. While one member of the family, *Torpedo* ClC-O, has been reported to have two channels, one per subunit, others are believed to have just one.

All functionally characterized members of the ClC family transport chloride, some in a voltage-regulated process. These channels serve a variety physiological functions (cell volume regulation; membrane potential stabilization; signal transduction; transepithelial transport, etc.). Different homologues in humans exhibit differing anion selectivities, i.e., ClC4 and ClC5 share a $NO_3^- > Cl^- > Br^- > I^-$ conductance sequence, while ClC3 has an $I^- > Cl^-$ selectivity. The ClC4 and ClC5 channels and others exhibit outward rectifying currents with currents only at voltages more positive than +20mV.

Animal Inward Rectifier K^+ Channel (IRK-C) Family

IRK channels possess the "minimal channel-forming structure" with only a P domain, characteristic of the channel proteins of the VIC family, and two flanking transmembrane spanners (Shuck, M.E., et al., (1994), J. Biol. Chem. 269: 24261-24270; Ashen, M.D., et al., (1995), Am. J. Physiol. 268: H506-H511; Salkoff, L. and T. Jegla (1995), Neuron 15: 489-492; Aguilar-Bryan, L., et al., (1998), Physiol. Rev. 78: 227-245; Ruknudin, A., et al., (1998), J. Biol. Chem. 273: 14165-14171). They may exist in the membrane as homo- or heterooligomers. They have a greater tendency to let K^+ flow into the cell than out. Voltage-dependence may be regulated by external K^+, by internal Mg^{2+}, by internal ATP and/or by G-proteins. The P domains of IRK channels exhibit limited sequence similarity to those of the VIC family, but this sequence similarity is insufficient to establish homology. Inward rectifiers play a role in setting cellular membrane potentials, and the closing of these channels upon depolarization permits the occurrence of long duration action potentials with a plateau phase. Inward rectifiers lack the intrinsic voltage sensing helices found in VIC family channels. In a few cases, those of Kir1.1a and Kir6.2, for example, direct interaction with a member of the ABC superfamily has been proposed to confer unique functional and regulatory properties to the heteromeric complex, including sensitivity to ATP. The SUR1 sulfonylurea receptor (spQ09428) is the ABC protein that regulates the Kir6.2 channel in response to ATP, and CFTR may regulate Kir1.1a. Mutations in SUR1 are the cause of familial persistent hyperinsulinemic hypoglycemia in infancy (PHHI), an autosomal recessive disorder characterized by unregulated insulin secretion in the pancreas.

ATP-gated Cation Channel (ACC) Family

Members of the ACC family (also called P2X receptors) respond to ATP, a functional neurotransmitter released by exocytosis from many types of neurons (North, R.A. (1996), Curr. Opin. Cell Biol. 8: 474-483; Soto, F., M. Garcia-Guzman and W. Stühmer (1997), J. Membr. Biol. 160: 91-100). They have been placed into seven groups ($P2X_1$ - $P2X_7$) based on their pharmacological properties. These channels, which function at neuron-neuron and neuron-smooth muscle junctions, may play roles in the control of blood pressure and pain sensation. They may also function in lymphocyte and platelet physiology. They are found only in animals.

The proteins of the ACC family are quite similar in sequence (>35% identity), but they possess 380-1000 amino acyl residues per subunit with variability in length localized primarily to the C-terminal domains. They possess two transmembrane spanners, one about 30-50 residues from their N-termini, the other near residues 320-340. The extracellular receptor domains between these two spanners (of about 270 residues) are well conserved with numerous conserved glycyl and cysteyl residues. The hydrophilic C-termini vary in length from 25 to 240 residues. They resemble the topologically similar epithelial Na^+ channel (ENaC) proteins in possessing (a)

N- and C-termini localized intracellularly, (b) two putative transmembrane spanners, (c) a large extracellular loop domain, and (d) many conserved extracellular cysteyl residues. ACC family members are, however, not demonstrably homologous with them. ACC channels are probably hetero- or homomultimers and transport small monovalent cations (Me^+). Some also transport Ca^{2+}; a few also transport small metabolites.

The Ryanodine-Inositol 1,4,5-triphosphate Receptor Ca^{2+} Channel (RIR-CaC) Family

Ryanodine (Ry)-sensitive and inositol 1,4,5-triphosphate (IP3)-sensitive Ca^{2+}-release channels function in the release of Ca^{2+} from intracellular storage sites in animal cells and thereby regulate various Ca^{2+} -dependent physiological processes (Hasan, G. et al., (1992) Development 116: 967-975; Michikawa, T., et al., (1994), J. Biol. Chem. 269: 9184-9189; Tunwell, R.E.A., (1996), Biochem. J. 318: 477-487; Lee, A.G. (1996) *Biomembranes,* Vol. 6, Transmembrane Receptors and Channels (A.G. Lee, ed.), JAI Press, Denver, CO., pp 291-326; Mikoshiba, K., et al., (1996) J. Biochem. Biomem. 6: 273-289). Ry receptors occur primarily in muscle cell sarcoplasmic reticular (SR) membranes, and IP3 receptors occur primarily in brain cell endoplasmic reticular (ER) membranes where they effect release of Ca^{2+} into the cytoplasm upon activation (opening) of the channel.

The Ry receptors are activated as a result of the activity of dihydropyridine-sensitive Ca^{2+} channels. The latter are members of the voltage-sensitive ion channel (VIC) family. Dihydropyridine-sensitive channels are present in the T-tubular systems of muscle tissues.

Ry receptors are homotetrameric complexes with each subunit exhibiting a molecular size of over 500,000 daltons (about 5,000 amino acyl residues). They possess C-terminal domains with six putative transmembrane a -helical spanners (TMSs). Putative pore-forming sequences occur between the fifth and sixth TMSs as suggested for members of the VIC family. The large N-terminal hydrophilic domains and the small C-terminal hydrophilic domains are localized to the cytoplasm. Low resolution 3-dimensional structural data are available. Mammals possess at least three isoforms which probably arose by gene duplication and divergence before divergence of the mammalian species. Homologues are present in humans and *Caenorabditis elegans.*

IP_3 receptors resemble Ry receptors in many respects. (1) They are homotetrameric complexes with each subunit exhibiting a molecular size of over 300,000 daltons (about 2,700 amino acyl residues). (2) They possess C-terminal channel domains that are homologous to those of the Ry receptors. (3) The channel domains possess six putative TMSs and a putative channel lining region between TMSs 5 and 6. (4) Both the large N-terminal domains and the smaller C-terminal tails face the cytoplasm. (5) They possess covalently linked carbohydrate on

extracytoplasmic loops of the channel domains. (6) They have three currently recognized isoforms (types 1, 2, and 3) in mammals which are subject to differential regulation and have different tissue distributions.

IP_3 receptors possess three domains: N-terminal IP_3-binding domains, central coupling or regulatory domains and C-terminal channel domains. Channels are activated by IP_3 binding, and like the Ry receptors, the activities of the IP_3 receptor channels are regulated by phosphorylation of the regulatory domains, catalyzed by various protein kinases. They predominate in the endoplasmic reticular membranes of various cell types in the brain but have also been found in the plasma membranes of some nerve cells derived from a variety of tissues.

The channel domains of the Ry and IP_3 receptors comprise a coherent family that in spite of apparent structural similarities, do not show appreciable sequence similarity of the proteins of the VIC family. The Ry receptors and the IP_3 receptors cluster separately on the RIR-CaC family tree. They both have homologues in *Drosophila.* Based on the phylogenetic tree for the family, the family probably evolved in the following sequence: (1) A gene duplication event occurred that gave rise to Ry and IP_3 receptors in invertebrates. (2) Vertebrates evolved from invertebrates. (3) The three isoforms of each receptor arose as a result of two distinct gene duplication events. (4) These isoforms were transmitted to mammals before divergence of the mammalian species.

The Organellar Chloride Channel (O-ClC) Family

Proteins of the O-ClC family are voltage-sensitive chloride channels found in intracellular membranes but not the plasma membranes of animal cells (Landry, D, et al., (1993), J. Biol. Chem. 268: 14948-14955; Valenzuela, Set al., (1997), J. Biol. Chem. 272: 12575-12582; and Duncan, R.R., et al., (1997), J. Biol. Chem. 272: 23880-23886).

They are found in human nuclear membranes, and the bovine protein targets to the microsomes, but not the plasma membrane, when expressed in *Xenopus laevis* oocytes. These proteins are thought to function in the regulation of the membrane potential and in transepithelial ion absorption and secretion in the kidney. They possess two putative transmembrane a-helical spanners (TMSs) with cytoplasmic N- and C-termini and a large luminal loop that may be glycosylated. The bovine protein is 437 amino acyl residues in length and has the two putative TMSs at positions 223-239 and 367-385. The human nuclear protein is much smaller (241 residues). A *C. elegans* homologue is 260 residues long.

The protein of the present invention is very similar to the dicarboxylate transporters. They bind a variety of divalent organic anions. Some of these carriers import acetylaspartate into the glial cells and play an important role in myelination. Others maintain succinate levels in

placenta and kidneys. Those expressed in the renal brush border may be relevant to pharmacological research. This sequence is also homologous to the family of sodium-sulfate transporters, which carry divalent inorganic anions across the cell membrane. Like its homologues, this transporter has 12 transmembrane helices.

Mitochondria and perhaps other organelles contain dicarboxylate transporters, which pump organic acids in and out of these compartments. Spatial distribution of divalent acids may affect the rates of the Krebs cycle, amino acid synthesis and other ergogenic and metabolic pathways. Sometimes, their local concentration exceeds physiological levels, which leads to formation of calcium stones.

The sequence presented here can be used to search for the specific interactors using affinity chromatography and the yeast two-hybrid system. Synthetic peptides and cytrate-derived compounds can be designed and used as inhibitors for these transporters.

For a review related to the dicarboxylate transporters, see references by Huang *et al., J Pharmacol Exp Ther* 2000 Oct;295(1):392-403, Chen *et al., J Biol Chem* 1998 Aug 14;273(33):20972-81, Pajor, *J Biol Chem* 1995 Mar 17;270(11):5779-85, Wang *et al., Am J Physiol Cell Physiol* 2000 May;278(5):C1019-30, Chen *et al., Arch Biochem Biophys* 2000 Jan 1;373(1):193-202.

Transporter proteins, particularly members of the sodium-dependent dicarboxylate transporter subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown transport proteins. The present invention advances the state of the art by providing a previously unidentified human transport proteins.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human transporter peptides and proteins that are related to the sodium-dependent dicarboxylate transporter subfamily, as well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate transporter activity in cells and tissues that express the transporter. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule or transcript sequence that encodes the transporter protein of the present invention. In addition structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen.

FIGURE 2 provides the predicted amino acid sequence of the transporter of the present invention. In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences that span the gene encoding the transporter protein of the present invention. In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. 55 SNPs, including 4 indels, have been identified in the gene encoding the transporter protein provided by the present invention and are given in Figure 3.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a transporter protein or part of a transporter protein and are related to the sodium-dependent dicarboxylate transporter subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human transporter peptides and proteins that are related to the sodium-dependent dicarboxylate transporter subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these transporter peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the transporter of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially

important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known transporter proteins of the sodium-dependent dicarboxylate transporter subfamily and the expression pattern observed Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known sodium-dependent dicarboxylate transporter family or subfamily of transporter proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the transporter family of proteins and are related to the sodium-dependent dicarboxylate transporter subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figures 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the transporter peptides of the present invention, transporter peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprising the amino acid sequences of the transporter peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the transporter peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated transporter peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. For example, a nucleic acid molecule encoding the transporter peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the transporter peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The transporter peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a transporter peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the transporter peptide. "Operatively linked" indicates that the transporter peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the transporter peptide.

In some uses, the fusion protein does not affect the activity of the transporter peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant transporter peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A transporter peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the transporter peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the transporter peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is

determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al., Nucleic Acids Res. 12(1)*:387 (1984)) (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the transporter peptides of the present invention as well as being encoded by the same genetic locus as the transporter peptide provided herein. As indicated by the data presented in Figure 3, the gene provided by the present invention encoding a novel transporter maps to public BAC AC034305, which is known to be located on human chromosome 17.

Allelic variants of a transporter peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by the same genetic locus as the transporter peptide provided herein. Genetic locus can readily be determined based on the genomic information

provided in Figure 3, such as the genomic sequence mapped to the reference human. As indicated by the data presented in Figure 3, the gene provided by the present invention encoding a novel transporter maps to public BAC AC034305, which is known to be located on human chromosome 17. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Paralogs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the transporter peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the transporter peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a transporter peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic

residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science 247*:1306-1310 (1990).

Variant transporter peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to transport ligand, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as transporter activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

The present invention further provides fragments of the transporter peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a transporter peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the transporter peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the transporter peptide, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide

fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in transporter peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol. 182*: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci. 663*:48-62 (1992)).

Accordingly, the transporter peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature transporter peptide is fused with another compound, such as a compound to increase the half-life of the transporter peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature transporter peptide, such as a leader or secretory sequence or a sequence for purification of the mature transporter peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a transporter-effector protein interaction or transporter-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

Substantial chemical and structural homology exists between the dicarboxylate transporter protein described herein and dicarboxylate transporters (see Figure 1). As discussed in the background, dicarboxylate transporters are known in the art to be involved in the major determinant of urinary excretion of citrate, the potent inhibitor of calcium salt crystallization urinary excretion of citrate, the potent inhibitor of calcium salt crystallization. Accordingly, the dicarboxylate transporter, and the encoding gene, provided by the present invention is useful for treating, preventing, and/or diagnosing dicarboxylate transporter related diseases such as kidney disorder.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, transporters isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the transporter. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Specifically, a virtual Northern blot shows expression in fetal liver and spleen. A large percentage of pharmaceutical agents are being developed that modulate the activity of transporter proteins,

particularly members of the sodium-dependent dicarboxylate transporter subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Such uses can readily be determined using the information provided herein, that known in the art and routine experimentation.

The transporter polypeptides (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to transporters that are related to members of the sodium-dependent dicarboxylate transporter subfamily. Such assays involve any of the known transporter functions or activities or properties useful for diagnosis and treatment of transporter-related conditions that are specific for the subfamily of transporters that the one of the present invention belongs to, particularly in cells and tissues that express the transporter. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Specifically, a virtual northern blot shows expression in fetal liver and spleen. In addition, PCR-based tissue screening panel indicates expression in human fetal liver.

The transporter polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the transporter, as a biopsy or expanded in cell culture. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the transporter protein.

The polypeptides can be used to identify compounds that modulate transporter activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the transporter. Both the transporters of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the transporter. These compounds can be further screened against a functional transporter to determine the effect of the compound on the transporter activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the transporter to a desired degree.

Further, the transporter polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the transporter protein and a molecule that normally interacts with the transporter protein, e.g. a substrate or a component of the signal pathway that the transporter protein normally interacts (for example, another transporter). Such assays typically

include the steps of combining the transporter protein with a candidate compound under conditions that allow the transporter protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the transporter protein and the target, such as any of the associated effects of signal transduction such as changes in membrane protential, protein phosphorylation, cAMP turnover, and adenylate cyclase activation, etc.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant transporters or appropriate fragments containing mutations that affect transporter function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) transporter activity. The assays typically involve an assay of events in the signal transduction pathway that indicate transporter activity. Thus, the transport of a ligand, change in cell membrane potential, activation of a protein, a change in the expression of genes that are up- or down-regulated in response to the transporter protein dependent signal cascade can be assayed.

Any of the biological or biochemical functions mediated by the transporter can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the transporter can be assayed. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Specifically, a virtual northern blot

shows expression in fetal liver and spleen. In addition, PCR-based tissue screening panel indicates expression in human fetal liver.

Binding and/or activating compounds can also be screened by using chimeric transporter proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a ligand-binding region can be used that interacts with a different ligand then that which is recognized by the native transporter. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the transporter is derived.

The transporter polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the transporter (e.g. binding partners and/or ligands). Thus, a compound is exposed to a transporter polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble transporter polypeptide is also added to the mixture. If the test compound interacts with the soluble transporter polypeptide, it decreases the amount of complex formed or activity from the transporter target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the transporter. Thus, the soluble polypeptide that competes with the target transporter region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the transporter protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of transporter-binding protein found in the

bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a transporter-binding protein and a candidate compound are incubated in the transporter protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the transporter protein target molecule, or which are reactive with transporter protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the transporters of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of transporter protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the transporter pathway, by treating cells or tissues that express the transporter. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. These methods of treatment include the steps of administering a modulator of transporter activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the transporter proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the transporter and are involved in transporter activity. Such transporter-binding proteins are also likely to be involved in the propagation of signals by the transporter proteins or transporter targets as, for example, downstream elements of a transporter-mediated signaling pathway. Alternatively, such transporter-binding proteins are likely to be transporter inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a transporter protein is fused

to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a transporter-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the transporter protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a transporter-modulating agent, an antisense transporter nucleic acid molecule, a transporter-specific antibody, or a transporter-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The transporter proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. The method involves contacting a biological sample with a compound capable of interacting with the transporter protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the

present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered transporter activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the transporter protein in which one or more of the transporter functions in one population is different from those in another population. The peptides thus allow a

target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and transporter activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Accordingly, methods for treatment include the use of the transporter protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or $F(ab')_2$, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the transporter proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or transporter/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Specifically, a virtual northern blot shows expression in fetal liver and spleen. In addition, PCR-based tissue screening panel indicates expression in human fetal liver. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or

abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the transporter peptide to a binding partner such as a ligand or protein binding partner. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a transporter peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the transporter peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a

protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the transporter peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the transporter proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. As indicated by the data presented in Figure 3, the gene provided by the present invention encoding a novel transporter maps to public BAC AC034305, which is known to be located on human chromosome 17.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridation conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. 55 SNPs, including 4 indels, have been identified in the gene encoding the transporter protein provided by the present invention and are given in Figure 3.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. As indicated by the data presented in Figure 3, the gene provided by the present invention encoding a novel transporter maps to public BAC AC034305, which is known to be located on human chromosome 17.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Specifically, a virtual northern blot shows expression in fetal liver and spleen. In addition, PCR-based tissue screening panel indicates expression in human fetal liver.

Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in transporter protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a transporter protein, such as by measuring a level of a transporter-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a transporter gene has been mutated. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Specifically, a virtual northern blot shows expression in fetal liver and spleen. In addition, PCR-based tissue screening panel indicates expression in human fetal liver.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate transporter nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the transporter gene, particularly biological and pathological processes that are mediated by the transporter in cells and tissues that express it. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. The method typically includes assaying the ability of the compound to modulate the expression of the transporter nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired transporter nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the transporter nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for transporter nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the transporter protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of transporter gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of transporter mRNA in the presence of the candidate compound is compared to the level of expression of transporter mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate transporter nucleic acid expression in cells and tissues that express the transporter. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Specifically, a virtual northern blot shows expression in fetal liver and spleen. In addition, PCR-based tissue screening panel indicates expression in human fetal liver. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for transporter nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the transporter nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the transporter gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative

compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in transporter nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in transporter genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the transporter gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the transporter gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a transporter protein.

Individuals carrying mutations in the transporter gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been identified in a gene encoding the transporter protein of the present invention.. 55 SNP variants were found, including 4 indels (indicated by a "-"). As indicated by the data presented in Figure 3, the gene provided by the present invention encoding a novel transporter maps to public BAC AC034305, which is known to be located on human chromosome 17. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point

mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a transporter gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant transporter gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985)); Cotton *et al.*, *PNAS 85*:4397 (1988); Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature 313*:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the transporter gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production

of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control transporter gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of transporter protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into transporter protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of transporter nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired transporter nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the transporter protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in transporter gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired transporter protein to treat the individual.

The invention also encompasses kits for detecting the presence of a transporter nucleic acid in a biological sample. Experimental data as provided in FIGURE 1 indicates expression in the fetal liver and spleen. Specifically, a virtual northern blot shows expression in fetal liver and spleen. In addition, PCR-based tissue screening panel indicates expression in human fetal liver. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting transporter nucleic acid in a biological sample; means for determining the amount of transporter nucleic acid in the sample; and means for comparing the amount of transporter nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect transporter protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by

reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the transporter proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the transporter gene of the present invention.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed

herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al., Techniques in Immunocytochemistry,* Academic Press, Orlando, FL Vol. 1 (1 982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology,* Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified transporter gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include

initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for

affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterotransporter. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene 67*:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene 69*:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology 185*:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res. 20*:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene 54*:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature 329*:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as transporters, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with transporters, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a transporter protein or peptide that can be further purified to produce desired amounts of transporter protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the transporter protein or transporter protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native transporter protein is useful for assaying compounds that stimulate or inhibit transporter protein function.

Host cells are also useful for identifying transporter protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant transporter protein

(for example, stimulating or inhibiting function) which may not be indicated by their effect on the native transporter protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a transporter protein and identifying and evaluating modulators of transporter protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the transporter protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the transporter protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS 89*:6232-6236 (1992). Another example of a

recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science 251*:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature 385*:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_o phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, transporter protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* transporter protein function, including ligand interaction, the effect of specific mutant transporter proteins on transporter protein function and ligand interaction, and the effect of chimeric transporter proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more transporter protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown in SEQ ID NO:2;

(b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3; and

(d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.

2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown in SEQ ID NO:2;

(b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3; and

(d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.

3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.

8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.

9. A host cell containing the vector of claim 8.

10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.

13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.

14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.

15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.

16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

18. A method for treating a disease or condition mediated by a human transporter protein, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human transporter peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO:2.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO:2.

22. An isolated nucleic acid molecule encoding a human transporter peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 or 3.

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 or 3.

```
   1 GTCTCCCTCC CGCGCGATGG CCTCGGCGCT GAGCTATGTC TCCAAGTTCA
  51 AGTCCTTCGT GATCTTGTTC GTCACCCCGC TCCTGCTGCT GCCACTCGTC
 101 ATTCTGATGC CCGCCAAGTT TGTCAGGTGT GCCTACGTCA TCATCCTCAT
 151 GGCCATTTAC TGGTGCACAG AAGTCATCCC TCTGGCTGTC ACCTCTCTCA
 201 TGCCTGTCTT GCTTTTCCCA CTCTTCCAGA TTCTGGACTC CAGGCAGGTG
 251 TGTGTCCAGT ACATGAAGGA CACCAACATG CTGTTCCTGG GCGGCCTCAT
 301 CGTGGCCGTG GCTGTGGAGC GCTGGAACCT GCACAAGAGG ATCGCCCTGC
 351 GCACGCTCCT CTGGGTGGGG GCCAAGCCTG CACGGCTGAT GCTGGGCTTC
 401 ATGGGCGTCA CAGCCCTCCT GTCCATGTGG ATCAGTAACA TGGCAACCAC
 451 GGCCATGATG GTGCCCATCG TGGAGGCCAT ATTGCAGCAG ATGGAAGCCA
 501 CAAGCGCAGC CACCGAGGCC GGCCTGGAGC TGGTGGACAA GGGCAAGGCC
 551 AAGGAGCTGC CAGGGAGTCA AGTGATTTTT GAAGGCCCCA CTCTGGGGCA
 601 GCAGGAAGAC CAAGAGCGGA AGAGGTTGTG TAAGGCCATG ACCCTGTGCA
 651 TCTGCTACGC GGCCAGCATC GGGGGCACCG CCACCCTGAC CGGGACGGGA
 701 CCCAACGTGG TGCTCCTGGG CCAGATGAAC GAGTTGTTTC CTGACAGCAA
 751 GGACCTCGTG AACTTTGCTT CCTGGTTTGC ATTTGCCTTT CCCAACATGC
 801 TGGTGATGCT GCTGTTCGCC TGGCTGTGGC TCCAGTTTGT TTACATGAGA
 851 TTCAATTTTA AAAAGTCCTG GGGCTGCGGG CTAGAGAGCA AGAAAAACGA
 901 GAAGGCTGCC CTCAAGGTGC TGCAGGAGGA GTACCGGAAG CTGGGGCCCT
 951 TGTCCTTCGC GGAGATCAAC GTGCTGATCT GCTTCTTCCT GCTGGTCATC
1001 CTGTGGTTCT CCCGAGACCC CGGCTTCATG CCCGGCTGGC TGACTGTTGC
1051 CTGGGTGGAG GGTGAGACAA AGTATGTCTC CGATGCCACT GTGGCCATCT
1101 TTGTGGCCAC CCTGCTATTC ATTGTGCTTT CACAGAAGCC CAAGTTTAAC
1151 TTCCGCAGCC AGACTGAGGA AGAAAGGAAA ACTCCATTTT ATCCCCCTCC
1201 CCTGCTGGAT TGGAAGGTAA CCCAGGAGAA AGTGCCCTGG GGCATCGTGC
1251 TGCTACTAGG GGGCGGATTT GCTCTGGCTA AAGGATCCGA GGCCTCGGGG
1301 CTGTCCGTGT GGATGGGGAA GCAGATGGAG CCCTTGCACG CAGTGCCCCC
1351 GGCAGCCATC ACCTTGATCT TGTCCTTGCT CGTTGCCGTG TTCACTGAGT
1401 GCACAAGCAA CGTGGCCACC ACCACCTTGT TCCTGCCCAT CTTTGCCTCC
1451 ATGTCTCGCT CCATCGGCCT CAATCCGCTG TACATCATGC TGCCCTGTAC
1501 CCTGAGTGCC TCCTTTGCCT TCATGTTGCC TGTGGCCACC CCTCCAAATG
1551 CCATCGTGTT CACCTATGGG CACCTCAAGG TTGCTGACAT GGTGAAAACA
1601 GGAGTCATAA TGAACATAAT TGGAGTCTTC TGTGTGTTTT TGGCTGTCAA
1651 CACCTGGGGA CGGGCCATAT TTGACTTGGA TCATTTCCCT GACTGGGCTA
1701 ATGTGACACA TATTGAGACT TAGGAAGAGC CACAAGACCA CACACACAGC
1751 CCTTACCCTC CTCAGGACTA CCGAACCTTC TGGCACACCT TGTACAGAGT
1801 TTTGGGGTTC ACACCCCAAA ATGACCCAAC GATGTCCACA CACCACCAAA
1851 ACCCAGCCAA TGGGCCACCT CTTCCTCCAA GCCCAGATGC AGAGATGGTC
1901 ATGGGCAGCT GGAGGGTAGG CTCAGAAATG AAGGGAACCC CTCAGTGGGC
1951 TGCTGGACCC ATCTTTCCCA AGCCTTGCCA TTATCTCTGT GAGGGAGGCC
2001 AGGTAGCCGA GGGATCAGGA TGCAGGCTGC TGTACCCGCT CTGCCTCAAG
2051 CATCCCCCAC ACAGGGCTCT GGTTTTCACT CGCTTCGTCC TAGATAGTTT
2101 AAATGGGAAT CAGATCCCCT GGTTGAGAGC TAAGACAACC ACCTACCAGT
2151 GCCCATGTCC CTTCCAGCTC ACCTTGAGCA GCCTCAGATC ATCTCTGTCA
2201 CTCTGGAAGG GACACCCCAG CCA (SEQ ID NO:1)
```

FEATURES:

5'UTR:	1 - 16
Start Codon:	17
Stop Codon:	1721
3'UTR:	1724

FIGURE 1A

HOMOLOGOUS PROTEINS:

Top 10 BLAST Hits:

gi|2811122|gb|AAB97879.1| (U87318) NaDC-2 [Xenopus laevis]	682	0.0
gi|4506979|ref|NP_003975.1| solute carrier family 13 (sodium-de...	629	e-179
gi|3065814|gb|AAC31165.1| (AF058714) sodium-dicarboxylate cotra...	626	e-178
gi|10280599|gb|AAG15426.1|AF201903_1 (AF201903) Na/dicarboxylat...	624	e-178
gi|2499524|sp|Q28615|NDC1_RABIT RENAL SODIUM/DICARBOXYLATE COTR...	602	e-171
gi|8132324|gb|AAF73251.1|AF154121_1 (AF154121) sodium-dependent...	543	e-153
gi|4322346|gb|AAD16019.1| (AF081825) sodium-dependent high-affi...	531	e-149
gi|5531902|gb|AAD44522.1|AF102261_1 (AF102261) sodium-dicarboxy...	515	e-145
gi|10439272|dbj|BAB15477.1| (AK026413) unnamed protein product ...	490	e-137
gi|2499526|sp|Q07782|NASU_RAT SODIUM/SULFATE COTRANSPORTER (NA(...	489	e-137
gi|9507109|ref|NP_062354.1| solute carrier family 13 (sodium/su...	486	e-136
gi|6912690|ref|NP_036582.1| sulfate transporter 1 >gi|6224691|g...	483	e-135
gi|2499525|sp|P70545|NDC2_RAT INTESTINAL SODIUM/DICARBOXYLATE C...	437	e-121
gi|6226757|sp|Q93655|YV06_CAEEL HYPOTHETICAL 66.2 KDA PROTEIN F...	400	e-110
gi|630683|pir||S43561 YCR37C homolog K08E5.2 - Caenorhabditis e...	390	e-107

EST:

gi|751038 /dataset=dbest /taxon=9606 /...	519	e-145
gi|2658836 /dataset=dbest /taxon=9606 ...	416	e-114

EXPRESSION INFORMATION FOR MODULATORY USE
gi|751038 Human fetal liver spleen
gi|2658836 Human fetal liver spleen

Tissue Expression: Human fetal liver

FIGURE 1B

```
  1 MASALSYVSK FKSFVILFVT PLLLLPLVIL MPAKFVRCAY VIILMAIYWC
 51 TEVIPLAVTS LMPVLLFPLF QILDSRQVCV QYMKDTNMLF LGGLIVAVAV
101 ERWNLHKRIA LRTLLWVGAK PARLMLGFMG VTALLSMWIS NMATTAMMVP
151 IVEAILQQME ATSAATEAGL ELVDKGKAKE LPGSQVIFEG PTLGQQEDQE
201 RKRLCKAMTL CICYAASIGG TATLTGTGPN VVLLGQMNEL FPDSKDLVNF
251 ASWFAFAFPN MLVMLLFAWL WLQFVYMRFN FKKSWGCGLE SKKNEKAALK
301 VLQEEYRKLG PLSFAEINVL ICFFLLVILW FSRDPGFMPG WLTVAWVEGE
351 TKYVSDATVA IFVATLLFIV LSQKPKFNFR SQTEEERKTP FYPPPLLDWK
401 VTQEKVPWGI VLLLGGGFAL AKGSEASGLS VWMGKQMEPL HAVPPAAITL
451 ILSLLVAVFT ECTSNVATTT LFLPIFASMS RSIGLNPLYI MLPCTLSASF
501 AFMLPVATPP NAIVFTYGHL KVADMVKTGV IMNIIGVFCV FLAVNTWGRA
551 IFDLDHFPDW ANVTHIET (SEQ ID NO:2)
```

FEATURES:
Functional domains and key regions:

```
[1] PDOC00001 PS00001 ASN_GLYCOSYLATIONN-glycosylation site
Number of matches: 2
      1    194-197 NSSL
      2    607-610 NVTH
--------------------------------------------------------------------------------[2]      PDOC00005
PS00005 PKC_PHOSPHO_SITEProtein kinase C phosphorylation site
Number of matches: 3
      1    222-224 THR
      2    336-338 SKK
      3    417-419 SQK
--------------------------------------------------------------------------------[3]      PDOC00006
PS00006 CK2_PHOSPHO_SITECasein kinase II phosphorylation site
Number of matches: 5
      1    222-225 THRE
      2    358-361 SFAE
      3    426-429 SQTE
      4    428-431 TEEE
      5    609-612 THIE
--------------------------------------------------------------------------------[4]      PDOC00008
PS00008 MYRISTYLN-myristoylation site
Number of matches: 7
      1     93-98 GLIVAV
      2    118-123 GAKPAR
      3    264-269 GGTATL
      4    271-276 GTGPNV
      5    460-465 GGGFAL
      6    468-473 GSEASG
      7    574-579 GVIMNI
--------------------------------------------------------------------------------[5]      PDOC00978
PS01271 NA_SULFATESodium:sulfate symporter family signature
          543-559 ASFAFMLPVATPPNAIV
```

FIGURE 2A

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainity
1	19	39	2.318	Certain
2	59	79	1.932	Certain
3	89	109	0.776	Putative
4	130	150	1.930	Certain
5	213	233	1.646	Certain
6	258	278	2.102	Certain
7	318	338	1.743	Certain
8	359	379	1.781	Certain
9	411	431	1.337	Certain
10	448	468	1.879	Certain
11	493	513	1.996	Certain
12	534	554	1.781	Certain

BLAST Alignment to Top Hit:
Alignment to top blast hit:

```
>gi|2811122|gb|AAB97879.1| (U87318) NaDC-2 [Xenopus laevis]
          Length = 622

 Score =  682 bits (1741), Expect = 0.0
 Identities = 332/619 (53%), Positives = 439/619 (70%), Gaps = 55/619 (8%)

Query: 1   MASALSYVSKFKSFVILFVTPLLLLPLVILMPAKFVRCAYVIILMAIYWCTEVIPLAVTS 60
           M S   ++   +++ I+F+ PL LLPL +++P K   C +VII+MA++WCTE +PLAVT+
Sbjct: 1   MVSIGKWILANRNYFIIFLVPLFLLPLPLVVPTKEASCGFVIIVMALFWCTEALPLAVTA 60

Query: 61  LMPVLLFPLFQILDSRQVCVQYMKDTNMLFLGGLIVAVAVERWNLHKRIALRTLLWVGAK 120
           L PVLLFP+  I+DS  VC QY+KDTNMLF+GGL+VA++VE+WNLHKRIALR LL VG K
Sbjct: 61  LFPVLLFPMMGIMDSTAVCSQYLKDTNMLFIGGLLVAISVEKWNLHKRIALRVLLIVGVK 120

Query: 121 PARLMLGFMGVTALLSMWISNMATTAMMVPIVEAILQQMEA-------------------- 161
           PA L+LGFM VTA LSMWISN ATTAMM+PI +A+++Q+ +
Sbjct: 121 PALLLLGFMVVTAFLSMWISNTATTAMMIPIAQAVMEQLHSSEGKVDERVEGNSNTQKNV 180

Query: 162 ------------------------TSAATEAGLELVDKGKAKELPGSQ-------VIFEGP 191
                                   T A    G E+ +K      P  Q       ++ E
Sbjct: 181 NGMENDMYESVMPSGKMALAIDNTYATENEGFEIQEKSTKDPEPSKQEKQSIGPIVIEPE 240

Query: 192 TLGQQEDQERKR---LCKAMTLCICYAASIGGTATLTGTGPNVVLLGQMNELFPDSKDLV 248
              Q E++++++   +CK M+LC+CY+ASIGG ATLTGT PN+V+ GQM+ELFP++ +++
Sbjct: 241 DEKQTEEKQKEKHLKICKGMSLCVCYSASIGGIATLTGTTPNLVMKGQMDELFPENNNII 300

Query: 249 NFASWFAFAFPNMLVMLLFAWLWLQFVYMRFNFKKSWGCG--LESKKNEKAALKVLQEEY 306
           NFASWF FAFP MLV+L  +WLWLQF+Y+  NFKK++GCG   E K+ EK A +V+  E+
Sbjct: 301 NFASWFGFAFPTMLVLLALSWLWLQFIYLGVNFKKNFGCGGNAEQKEKEKRAFRVISGEH 360

Query: 307 RKLGPLSFAEINVLICFFLLVILWFSRDPGFMPGWLTVAWVEGETKYVSDATVAIFVATL 366
           +KLG ++FAEI+VL+ F LLV+LWF+R+PGFMPGW T+++ +G  + V+DATVAIFV+ +
Sbjct: 361 KKLGSMTFAEISVLVLFILLVLLWFTREPGFMPGWATISFNKGGKEMVTDATVAIFVSLM 420

Query: 367 LFIVLSQKPKFNFRSQTEEERKTPF-YPPPLLDWKVTQEKVPWGIVLLLGGGFALAKGSE 425
           +F   S+ P F ++   +   K     PP LLDWK   EK+PW IV+LLGGGFALAKGSE
Sbjct: 421 MFFFPSELPSFKYQDTDKPGMKPKLRVPPALLDWKTVNEKMPWNIVILLGGGFALAKGSE 480

Query: 426 ASGLSVWMGKQMEPLHAVPPAAITLILSLLVAVFTECTSNVATTTLFLPIFASMSRSIGL 485
            SGLS+W+G+++ PL ++PPAAI LIL LLVA FTECTSNVATTTLFLPI ASM+++I L
Sbjct: 481 ESGLSLWLGEKLTPLQSIPPAAIALILCLLVATFTECTSNVATTTLFLPILASMAKAIQL 540

Query: 486 NPLYIMLPCTLSASFAFMLPVATPPNAIVFTYGHLKVADMVKTGVIMNIIGVFCVFLAVN 545
           NPLYIMLPCTLSAS AFMLPVATPPNAI F+YG LKV DM K G+++NI+GV  + LA+N
Sbjct: 541 NPLYIMLPCTLSASLAFMLPVATPPNAIAFSYGQLKVIDMAKAGLLLNILGVLTITLAIN 600

Query: 546 TWGRAIFDLDHFPDWANVT 564
           +WG  +F+L  FP WAN T
Sbjct: 601 SWGFYMFNLGTFPSWANAT 619  (SEQ ID NO:4)
```

Hmmer search results (Pfam):
No match

FIGURE 2B

```
   1 TTCAACCATT GTGGAAGACA CTGTGGCGAT TCCTCAAGGA TCTAGAACCA
  51 GAAATATCAT TTGACCCAGC AATTTTATTA CTGGGTATAT ACCCAAAGGA
 101 TTATAAATCA TGCTGCTATA AAGACACATG CACACTATTT ACAATAGCAA
 151 AGACTTAAAA CCAACCCAAA TGTCCATCAA TGATAGACTG GATAAAGAAA
 201 ATGTGGCACA TACATACCAT GGAATACTAT GCAGCCATTA AAAATAATGA
 251 GGTCATGTCC TTTGCAGGGA CATGGATGAA GCTGGAAGCC ATCATTCTCA
 301 GCAAACTAAC ACAGGAACAG AAAACCAAAC ACCACATGTT CTCAGTCATA
 351 AGTGGGAGTT GAACAGTGAG AACGCATTGA CACAGGGAGG GGAACATCAC
 401 ACACGGGGGC CTGTCAGGGG GTTGGAGGGC AAGGGGAGGG AGAGCATTAG
 451 GACAAATACC TAATGCATGT GGGTCTTAAA ACCTAAATGT CCGGTTGATA
 501 GCTGCAGCAA ACCACCATGG CACATGTATA CCTATGTAAC AAACCTGCAC
 551 ATTCTGCACA TGTATCCCAG AACTTAAAGT AAAATTAAAA AAAAAGAAAA
 601 GAAAAAAGAA CTGAAGTTGT TTACTTGCTC TCATTCATGC ATCCCGGAGA
 651 AAAAGGTTTG AGTGCACATC CTGGATTAGG CACTGAGAAA GGCACTAGCT
 701 GGACAGGTGG TGATGAATAA AACAGACAGT AAATAGAAAT TACATCATAA
 751 TAATGTGTCA TATATTTTAA AATAGCTACA AGATATTTTA AATGTTCTCA
 801 CCACAAAGAA ATGACAAATA TTTGGGCCAG ACGCGGTGGC TCACGCCTGT
 851 AATCCCAGCA CTTTGGGAGA CCGAGGTGGG CGGATCACCT GAGGTCAGGA
 901 GTTCGAGACC AGCCTGGCTA ACATGGTGAA ACCCCATTTC TACTAAAAAT
 951 GCAAAAAATT AGCCGGGCGT GGTGGTGCAC ACCTGTAATC CCAGCTACTT
1001 GGGAGGCTGA AGCAGGAGAT TTGCTTGAAC CTAGGTGGCA GAGGTTGCAG
1051 TGAGCCGAGA TCGTGCCACT GCACTCCAGC CTGGGTGACA GGAGCACAAC
1101 TCTGTCTCAA ACAAACAAAC AAAAAACAAA AACAAGAGAA ATGATAAATA
1151 TTCGAGTGAT AAATATGCTC ATTAGCCTGA TTTGAACACA CCACAATTAT
1201 ACACACATTG AAAAATCACA TGGTACCCCG TAAATATAGA CAATGATTTG
1251 TCAATTAAAA ATGAAATAAC ACTTAAAAAA TAAAAAAGTA AAAAGTAAAA
1301 ATTACACCAA TAAATATAAG AGGTACAAAT TGTGCTAAGT GCCCTGGGGA
1351 CACAGGAAAG GCGGGAAAAC CCAGGGCTAT ATGCATGAGA GTTACAAAGG
1401 GAAAAGGACA GGAGGGAGGC AATTGCAGGA GGGGCTTGGG AGAATGCATG
1451 TCCTTGGGTG CAGGTCACAG GAAGGAACTC ATGAGCTTGA TTCAGGATGT
1501 GTTGAATTTT CGGGCCGAGA CACGTCCAGT CTGCGGAAGG CTGGACATCT
1551 GGGACTCTGG CATCATGGCT GGGTTGAAGG CAGAGGATGG TAATCACTAA
1601 GGAGCCGGCT GTGGTTAGGC CACCAGCATG GATGAGACTC CCCAAAAGGA
1651 AGCTGCAGAA TGAGAGGCAG GCAGAGGAGA GGAAAGAAGA AAATCACAGA
1701 GGTGGGGATG TCTTTGCATC CGTGTGTCTC CAGTGCCCAA AACAGGGCCT
1751 CGCGGAGAAG AGGTGCTCGG CACCTGTCTG TTGCCTGGCG GGCTGAATGA
1801 ATACATGGGC GACTGTCTCA GTGTCGCCTT AGTTGTGTCC CTTCCTCTCT
1851 AGAGCTCCGT TTCCCTCTGA CCTGGGTCGG GCGGGCAGCT GCGGCTGCTG
1901 AGGCTCGGTG GGGCCCCTCC AAGACGCGTG TCCGCATCTG CCCGCCGGGC
1951 GTCTGCGGGG TGCAGCGTCC ACTGGAGCGC GACAGCCCCT GGGACAGAGG
2001 AGGACAGTGG CCTCGCTTCC CTGTGCGATC GCCCAGGAGC TCCGGGCCGG
2051 AGAGTGCGAG CGGGGAAAAG GGGTCCTGCA CCTAGAGTGG GGCGGACGTG
2101 GCGAGGAAGC CAGGGGGGAC CGGGAAGCGA GGCCCGCGGT GCGGAGGGCG
2151 CGGGGCGTGG GGGGACACCT CTCGGAGAGA CACCGGAGGG GCGGAAGTAA
2201 GGAGATGGAA AGGAGAGGGA GATCGGGGAG ATAGACCTGA GAGACCCAGA
2251 GGCCTGCAGA GAGTTTCATC CGGGACCCTT CAGAGCCCAG GAAAGAGCAG
2301 ATGCGGACGC GGGAGGGCGC CTTACGCCAA AGCGGGCAGC ACCAGTGACC
2351 AAAACACGCC CCGCTTGGCA GCCCCGGGAC GCACCTCTGC CTCGGCAGCG
2401 CAGGAGAGGC TTGGACAGCG CGAGATGCTA GGGCCCAGGC TGCCCCTAGA
2451 GGGCTGGCCC GAAGCGTTGG AGTCCAAAGA CGCCTCCCAC CGCCGCCGGG
2501 TGGCAGAATT GGGGGCAGGC GCGTCCCACA GACCCCGAGG GGTGGCCCCG
2551 CCCCAGGGCC GCGGGGAGGC GCCCCCGTGC GGGGCGGAGT TGTCACCGCC
2601 CCCTCCCCAA TCCCCGGGGA CTGTGGCCCC TTCTTAAGCC CGCGGCGCCT
2651 CTAGCTGCCC CTCACTCGTC TCGCCCGCCA GTCTCCCTCC CGCGCGATGG
2701 CCTCGGCGCT GAGCTATGTC TCCAAGTTCA AGTCCTTCGT GATCTTGTTC
2751 GTCACCCCGC TCCTGCTGCT GCCACTCGTC ATTCTGATGC CCGCCAAGGT
2801 CAGTTGCATC TCCAGGCAGC CCTTCGGACA CCCGGCGTCC TGTGCCCACT
2851 AACGGGCACC GATCCCGGGA GCCCTGAGCT GGAGCGCACG GATTTCGCGG
2901 GGAGCACAGC TCTCCCGGGG CGCGCGCACT CAGGAGCTCC AGGTGCCGGA
2951 TGGGAGGTGC CCTGTAAAGA ATCTGAGGGG CATGGCGACC CCAGGGCGCA
3001 CCACCCTTGG GGTTTACAGA TCCCAGGGCG CAAGAGCCGT CCAGGCAAGC
3051 ACGGAAACCT CGAAGTGAGC ACAGATCTCA GCCACACAGA TCCCAGCCTT
3101 AGGCTCAGCC CCTGGCTCCG AATCGAATCT CCCACAGTGC ATAACCCTGT
3151 TTCCCCCCAA AATGCCACCT GCGCCAACAG GGAACCTGGG AGCTTGCCTT
3201 TCCCTCTCTC TCTCCTGTCT TTTCCCTTCG CCAAAGAAGA CTTCAAGCTG
3251 TAGGTGGCTT CTGCCGTCAG GAGGGACCTA CAGGAAAAAA ATCATCACCC
3301 ACGTGGATCC TGCGCTGTCT TTGCCACTCT CTGGCCCTTC CTTGGGCCTT
3351 AGTGTCTCTA TCTATGATCC ACATTCCTTC CAACCTGGAG AGCCACATCT
3401 GATTCATAAT CCTGCTCCAG GCTGCAGGCA GGTTGGGGTT GGCCTGCTTT
3451 GCCTCCTGCC TGCTGGGGCT GTAGCAGGAG GCGGGACACA TTCCCAGAGC
3501 TCGCAGCCTT GGGTGGCAGG ACCTGGAGTT GCAGGGAAGC TTCCTCCCAG
3551 GCCCTAGTCT CCTAATGCTT CTGTGAGGGA GAGAGAGAAT GAATGGCCTT
3601 GGCCGCAGGG TGGGCGCAGG CTCCACTGGG CTGTGCACAG CCAGTTTGGC
3651 GGAGGCCCAA GCCCTTTGAA GCCTTTTGTG GCTGCTGGCT GCTCCTTCCT
3701 CGTTCCTTCT TTCAGCCCTT TCACTCTCAG CCCAGACAGG AAACCTCCAG
```

FIGURE 3A

```
3751 CTCCCCACCT CCCCTCCCCA GGCAGGTTTG GGAAACAGAG GAGCTCTTCA
3801 GGGGATGCTC TGGGGGGGAG CTCTAGAGGA AGGGAGTGCA CTGGGGTGTC
3851 AGGAAGCCAA CCTGCAAGAG AACTGGACTT CCACCATTCT GTCATCTGTG
3901 TGACTTTAGC CAAGTGCCTG TGCTCTCTGG GCCTTGGTGT TCTCATCTGT
3951 ACAATGGGGC TAGGTGGTTT GGTTCTGACA CTTTAAGGCT TTGGGAATCA
4001 AAAGGGATTA ACGTTAGTTC CTAGGATGGG GGAGGGGGAG ACTGGGAGGA
4051 GCCCTTGGGT GGGCCTAGCA CAGGCCCTGG ATGGGTCAAG GACAGCAGAT
4101 CCAACTGTGG AGGCTGTGCA GTTGCCACAC CAACTGTGGG CAGCCACACT
4151 TCCTCTGTAG CAAGAAGTCT GGGGTTGTTA TTGTCCAGGG GAAGTAGCCA
4201 GGCAGGAGAG CTGCGATTTC AGCTCTGCTG GTAGGGAGTG ATGTTCCCTG
4251 GAATGATTAT AGTAGCTTGG CTGACCTTCC TGCCACAGGA GACCCCACTC
4301 ACACACACAC ACATACACAC TCTCTCTCTC ACACACACAC ACTCACAAAC
4351 ACACACAGAC ACACACACAA ACACACAGAC ACACACAAAA CACACGCACA
4401 AACACAAACA CACACAAAAA CACACGCACA AACACACACC CAAACACACA
4451 AACACAGACA CACAAACACA CACACAAACA CACACACAAA ACACACAAAA
4501 AACACAAACA CACACATACA CATACACACA CACACACACA CACACCCTGA
4551 ACTGGAAACC CTAACTCAGT GTGTGTGTAT GTGTGTGTAT GTGTGTGTGT
4601 GTAAGAGAGA GAGAGAGAGA GATTAAGCTG TCCTTTGAGT GAGGACCAGG
4651 GAGGGGAAGA AGAGAACCCA GGGAGAGTCC TTCCAAAGGC TGCCTTCACG
4701 AGCTTTCCTT CTGGCGGGGT TGGGTGAGGA CCCTGGACCT TGTCTTCTTG
4751 TTTTTTCCCT TTCTGCCTGT TTTGGTCACC CTGCCCCCAC CCTCCATGGC
4801 CGCCCCATTG TGCAAGGAAA CCCAGAGGGT ACACAGCACG GGCAGGGCAG
4851 CTGGGAAGCT GGTGAGAAGC TGGGAGGACC TTGGCAGCCT GAGCAACACA
4901 GTCCTTGCCA GGAGGTGACT CCCAGGGCAC GCCACCCTCT GCCAACACCC
4951 AGGCCTCTCT CCTCACCGAC TGTCTCCAGT TTTCCTGTCT CCACCTGGAT
5001 TCCCTCCTGG CCTCATCTCT GCTCCACTCT CTCTATCCTT CCTCTGGGTC
5051 TTTTTTTAAT TGAAAAAAAA TTTAATGAAA TAAATGATAG ATTTCTTGTA
5101 TCACTTATTT TATTAAAATG TAAAAGGTTT CTTTTTTGCA AATCTGTAAG
5151 ATATAAAGTA AAAATAAAAG TACACTCAAA TCCCATAAGT TATTCACATT
5201 TTGATGAACA TCTTTCCAGA TGAATCTCTC TCTCTCTTCC CAGACACACA
5251 CACACACACA CACACACAGT AGGTTTTGCC TGCATTTTTT CATTAAGTGG
5301 TGTGTCAGGA CACCCTGCCT TGTTAATGTT GAACTTTTCT AACATCCGCT
5351 TCCCATCCTG CCCTCTCCCT TGACACTGTG GAGGCATTCT AGACTAGGGG
5401 GGTCAAGCCT GTTGACTTCA GGGATGAGGC ACCTCCTGGG CTTCTAAATA
5451 GTGGCGCGGA GGTGAGGGGG CAGTTAACCT TGTGTCTCGT CCTCTTTCCT
5501 AGTGGGTCTG CTTGACTCCT CCAGGAACGC ACAGTGTACA TTGGTGACGC
5551 ACGCCACCTA CTGCTTCTAA GTTTAGAGAA TCAAAAGTTA CCGAGGACTT
5601 TGTGCGCCAT ATGGGAAGAA TGAGCACTCT TAAATCCACG ATTTGCAGAT
5651 GAAGACATGA AACAAGAGGG GACAGGGACC AGGATTGGGA GCAGGAGGAG
5701 TAATTTATGA GCGACATTGT TTAGAATTGC TATCACTTGA TGATAGTAAG
5751 AAGCAAACTA ATTTTTAGCT AATATTATTG TTTTAAAATT CTCTCCAATG
5801 CGCCCTCTCA TTGTCTGCCC CTGGAGGCAT CATTCTGATG GCCTGCCCAG
5851 GGTACACCCC GACACCAAGC CCCAAGGAAG TTAGTGGCTG CCAAAGGCCA
5901 GACAGTGGCT GACAGTGGGC GCCAATCATA TCTGTCTGGT GTCAAAGCCT
5951 GGGCTTCCAG TCACGCTGCT GTTCCGCCTT TAGTTCAGCG GCTGTCAACC
6001 AGCATTGACC CTTTCTCCTG CCCTCACCCT GCCCCACAAC AGGGGAACTT
6051 TGGCAAAGTA CAGAGACATT TTTTGCTGTC CAACCTGGAG ACAGTCTTAC
6101 TGGCATCTCA TAGGTGGAGG CCAGCGGTGC TCTAAACACC CTGCAGTGCA
6151 CAGCTCCCAC AACAAAGCAT CATTTAGCCC AAAATGTCAG TGTGCCGAGG
6201 CTGAGGGACC CTGCCTCCCA GTAGGGAGGT GCCCTGGTTT GCTCGTGGGA
6251 TGCTGAAAAA AGATTTATTT TTTTGTGGCT GATAACACAA CCCTGACAAA
6301 GAATTTCCAA GTCTTCCTGC ACTGTTTTGT GCAAATAATA CATACGCTCT
6351 TCTGGGTGAT GAGAAGCAGG GATTGTGTAC AGGTGCATCT GTTCTTCAGC
6401 AGCATTGTCA GAGTTAAACT CAGATGAATG CTATTGATTC CTTTAATAAA
6451 CATTTGCAAA GATGGCCGGG CACAGTGGCT CATGCCTGTA ATCCCAGCAC
6501 TTTGGGAGGC CGAGACAGGT GGATCACGAG GCCAGGAGAT CAAGACCATC
6551 CTGGCCAACA TGGTGAAATC CCCTCTCTAC TAAAAATACA AAAATTAGCC
6601 GGGCGTGGTG GCGCTTGTCT GTAGTCCCAG CTACTCAGGA GACTGAGGCA
6651 GGAGAATCGC TTGAACCTGG GAGGTAGAGG CTGCAGTGAG CCAAGATTGC
6701 ACCACTGCAC TCCAGCCTGG GGACAGAGCA AGGCTCTGTC TCAAAAATAA
6751 GTAAGTAAGT AAATAAATAA ATAAATAAAT AAATAAGCAA GAATTTGCAA
6801 AGATATCCTA AGTGTTGGGC CTGTTCTGGA TGCTGAGGAC GGTGATCTAC
6851 AAATACAGCA GGTTCTTGAA TAATGTTGAT TCATTCAATA TCATTTCATT
6901 ATAATGTTGA TGAGGGAGGG AAAAAAAAGG AAGGATCCCT TGAGCCCAGG
6951 AGATGGAGGT TACAGTGAGC TGTGACCGTG CCACTTCACT CCCACCTGGG
7001 CAACAGAGCC AGACCCTGTC TCAAAAAAAA AAAAAAAGAA ATAAAAGAGC
7051 GAGAGAGAAA GAAAAGAAAA TGATTACTGG CTGGGGCCAC TGTCTGTGTG
7101 GAGCGTGCAC ATTACCCTCG TGTCCACATG GCTTTTCTTT GGCTAGTATG
7151 GTTTCCTTCC ACATCCCAAA CCCGTGCACG TTAGGTGAAT TGGAGTGTCT
7201 GTATGGTCCC TGTCTGAGTG AGCGTGGGCG TGCGTGTCAG TGTGCATTCT
7251 GCAATGGGAT GGCATCTTGT CCAGGGCTGG TTTCCACCTT GTACCCTGAG
7301 CTGCCGGGAC AGGATCTGGT CACCCAAGAC CCTGACCTGC TGTAACTGGG
7351 TAAATAATTA TCTAACTTGT TTTCAATGTT TCTTAAGTAT ATGTATAGCT
7401 CACATTCCCT TCAGTGTTTA ATATTGGAAG TGTTTTGGTC TTTATTTAAG
7451 ATCTCCGTGA TGTTTTTGTG ACCAGAAATA TGCTGTAGAA ATTTAACTGT
7501 TGTTTCTAGC AATTTGCCTA TGGGAATATT GGCTTATGTT GTTTCGCTTA
```

FIGURE 3B

```
 7551 CGCATTGCAA TTTCCAAAAA CCAATCAATG ATGTTAAGTG AGGACTCACT
 7601 GTACTGTTTG TGCTTTCGAG TCACGCACTG GTTGTGGTGG TAGAAGGACA
 7651 GTTGAGGAAA CAGTGACAAC TCCATATGCT AATGGCTGGG GAGGGTACTC
 7701 AGAGGAAGGG CACAAACCAG ACTATAGAAG AGGCGCAGGG AGACATCTAA
 7751 GAAGGAACTC TGAGGTTGGG CGCGGTGGCT CACGCCTGTA ATCCCAGCAC
 7801 TTTAGGGTGC TGAGGTGGGC GGATCATGAG GTCAGGAGTT CGAGATCAGC
 7851 CTGGCCAATG TGGCAAAACC CCGTCTCTAC TAAAAATACA AAAATTAGCA
 7901 GGGCGTGGTG GCAGGTGCCT GCAATCCCAA CTCCGGAGGT TGAGGCAGGA
 7951 GAATCGCTTG AACCCGGGAG GTGGAGGTTG CAGTGAGCTG AGATTGTGCC
 8001 ATTGCACTCC AGCCTGGGCA ACAGGATCGA AACTCTGTCA CACACACACA
 8051 CACAAAAAAT ACTGATGAAA CATAAAACAA CCTAGGGAGG TGGCTAGTTT
 8101 TATCACATAA TTATTATTAC TTTTATTTCA ATAGCTTTAG GGGTACACGT
 8151 AGTTTTCGGT TACATGAATG AATTGGATAG TGGTGAAGTC TGAGATTTTA
 8201 ATCCCTCCCT CCTATCCCAC CCTGTCTGCT TCTAAGTCTC CAGTATCCAT
 8251 TCGACCACGC TATATACCTC TGGATACCCA TAGCTTAGCT CCCACTTATA
 8301 AGGGAGAACA TGCACTATTT GGCTTTCCAT TGCTGAGTCA CTTCTCTTAG
 8351 AACAATGGCC TCCTAGGCGG CAAGAGCGAC ACTCCATCTC AAAAATAAAA
 8401 TAATAATAAA ACCAAAAAAA CCAGGTATTT TATTCTTCTT CTCCTTCTCC
 8451 TCTTCCTCCT TTCTTTTCTT CCTTCTCCAT CCCCCTTCCT CTCTTTCTTC
 8501 TATCCCCTCC TCCTCCTTCT CCTCCTTCTC CTTCTTTCTC CTTTTCTTCC
 8551 TTGTCCTATT CTTGATCTTT TCTTTTGAGA GGCAGCTAAT CCAAGGTTTG
 8601 AGAAGATGAA AGAACGTGCC TAGAACCACA CAGCTGGGAA GGAGGGAGGC
 8651 AGGGAGGAGG GGTGGGAATG GGGCAGGAGT CCTTTGCGAA TAGATCCCTG
 8701 GCCTGACCCG GGAAAGCTGT GCTGACCAGG GCTGGGGAAC AAGATGACTT
 8751 TGAGGGGAAT CCCTCTGAGA TCAGCACTGT GTCTTGACAA TCCATGCCAG
 8801 CCGCCGTCCG GAGTGTTCTG GGGGTGGGGA GAGGGAGGCG GCAACACGCT
 8851 GAGGCCTCAG GACTGTCTCT TCAGTTTGTC AGGTGTGCCT ACGTCATCAT
 8901 CCTCATGGCC ATTTACTGGT GCACAGAAGT CATCCCTCTG GCTGTCACCT
 8951 CTCTCATGCC TGTCTTGCTT TTCCCACTCT TCCAGATTCT GGACTCCAGG
 9001 CAGGTGAGCA GACCCAAGGG ATCCTGGTGA CTTTCTGGTT CTCCCCTTCT
 9051 CTCTTTCTCT AGTCCCCACT GTGAGTCGCA CAGGCCTGGG GGTGACCGGA
 9101 AAACCCTCAT TTGTGGATTC TCCCTGGCAG GGAGACACCA CTCGAGCCTG
 9151 CATCCCCACT CCAAGCTGTC CCTGAAGTCA GCATCTGGGG ACTGGGTGGC
 9201 TCTAGTGTGT GGCAAGGGAC AGTCCTGATG AGGCCTTCGT GCCACGCTCC
 9251 AGGTGTGTGT CCAGTACATG AAGGACACCA ACATGCTGTT CCTGGGCGGC
 9301 CTCATCGTGG CCGTGGCTGT GGAGCGCTGG AACCTGCACA AGAGGATCGC
 9351 CCTGCGCACG CTCCTCTGGG TGGGGGCCAA GCCTGCACGG TAATTACGCC
 9401 TTCTCTCTCT TGCCACGTGG CTCTGCATGA GCCCCAGGGC TGGAAGGGGG
 9451 TGGAGGATGG CACAGACCAG GCCATCCACT GGTGAGGGCT GGCCATGGGC
 9501 TTACCTGGAC TTGGCTGGGT GGGGTGCAGT TATAGCTTTA GTGGGAGAGA
 9551 CCAGATGCAT GCGTGGTGGT GGCACATGGT GAGCAGCAGT AAGTAAGGGT
 9601 CCTCGAATCC AGAGGAGGTG GGTCAGCAAG AGTCCTTGCA GGCTTGGAAG
 9651 GCTTTCTGGG GGAGGCAGCT AGCTGCAGGG TTCCACCGGG AACAAATTGG
 9701 ATAGAGGCTG GATCAAGCTG TGTCTGATAG GATAAGGGAA GCAGGCCAGA
 9751 AGTGGCTCAA CTACCCAGCT CATGGGGAAG CAGAAAGGTC CTCTCTCCAA
 9801 GCTGGAGCAT CTATTCCCAC TGCAAAGAAG CTTCTTATCT TCCCCGATAT
 9851 CACTCAGTAC CCCAGCTTCT CTCTCCATTT CCAGGATCTC TCCTGCCAAT
 9901 CTAGCTAGCC ATTTCCAGCT AAGCCATGGA GTCAATATAA TCATAATCAT
 9951 AACCATAATC AATCATGATC ATAATGGGTA TATTGAGTGT CTACAAGGCC
10001 CCAGGCATGA TACCAGGAGC TTATGAATTG CCTCATTTAA TTCTTACCAC
10051 AACTCTGAGA GCTGAGTATT CTTACTGCCC ACATTCTGTG GATGAGGGAT
10101 TGGAGGCAGA GAGGGATAAA GTGATTTGCT CATGGACACA CGGGGATTGG
10151 ACCCAGCTTC TCTGATAAGG CCTGTGTCCT CTCTAATCAG AAACTCAGGG
10201 CATATCTTCC TTTTGAGACA ATGTGTCCCC TCAATGATGG CACGTCCTTG
10251 GCCCAGCCAT CAGGAGTCAG CTGCTGGTCA GTTTACGGTA AATTCCTCCT
10301 GAGGCGCCCC TGTGTCAGGG GCTGTGCTAG ACCCTGAGCA CACACAGACG
10351 CTAGCCTGCC CTGCAGGCAA CCCATGCCCG GGGCTCCCTT GGTGCCTCAG
10401 ATTCCCCTGA GCAGGGGACT TTGTGGCTCT CTGATCTGTT ACACATCCTG
10451 GCATTGATTC CTTCCAGTAA TTGCTTTAGC ATCAAATCAA AAGCCATCAT
10501 ATTTTCTAGA AATGAGAGAC CCCAGGAAAG TGGACCTCAG GGCCCTCAGA
10551 ATTCTTCTGC TTGGCTCCCT TGAGTGGCCA GCTTGGGTGG GAGGCCACTC
10601 CAGTGGGTTT CATTCTGCAG CATGCTGGAG AGCTTCCACT TCCAAACCCA
10651 AGTTCACACA TGCTTCTGTA TCCTTCCTGC CACCTTGCTC CTCTGAGTAT
10701 GGTCTCCGGT TGTCCAAGGC ACTGCCTGTC CTGGGAGTCA CCTGTATGTG
10751 AGGCACCCTT GGTGCCTTGA GATATCATGT AGAAGCCTTG GTTCTTCTCA
10801 GACAACTCCA TTCATGCAAA CTCTCCCCCT CCTCCTAGCC TGGGTCCCGG
10851 GCTTTGTTTT TTTTTGGGTC CATAATGTCT GCCTGTGTGG ACAGCAGCTT
10901 GGGCCCTGGT GCAGAACAGC TCCTAGGTCC CTTCTTCAGG CTCCTACCCC
10951 TGCCCCTGCT CCTACCCCCA GGTGAATTAG GAGCCCTGAG GAGGAGCCTG
11001 GCTGCAGCGA GGCCCACAGA CTGAGAGTAG CTGAGCTCCT TCTGTCCCTA
11051 GCCTTGGACA GCTGGGGCAT GTAGAGCCAC AGAGCAGAGT CAGGCCCTGC
11101 CCTGCTCACA GCCCAAGGAG AGAGCAGACA TGGAAACAGG TGCTTTGAAC
11151 CCAGCACAGC GATGATTAGA GTAGGGGGAA GGATTGAGAA GGGTCAGGCC
11201 AGCCCCACCT GGTGCACACA CTGAGAGCGT GGTCCCAGAG GAGGGATGTT
11251 GTTTGAGCAG GCTCTGAAGG ACCATGAGGA GTCTTCCTGA TAGACAGCAG
11301 AGAAGGGAGC AGGGGTTACA AGCAAAGGGA GTGTTTCTTC TGAATACTGT
```

FIGURE 3C

11351 TTGTGTGATA GCTCCACTGC AGCATGGAGG GGTCAAAGTG TATGTGCGGG
11401 GCGGAGGGGA GATGGCAGGG TTGGAGTGGC AGCCGGGAGA ATAGTCACAC
11451 TTTCCCAAGC TCCCTCCCCA GCTCACCCTA CCCCTACTCT GCTTAGCCCT
11501 TCTGAACTTC TGAGAGGTGC AACAGAGTTT GGGGGTGGGT GGGAATTTCC
11551 TAGCCAGAAG TGGGAAGCTG GGGCTGCCTG CACATAGGGG TATTCCAGCA
11601 CACCCTAGGG CAAGCTCATA TTGAGTTGGC ACCATCTGGA TGCCTGGGCT
11651 TCCCCTGCTA GATGGTGGGG CAGGGGTGCT CCTTAGAACC ACGACTGGAT
11701 CTGAGGCCTC TTGGTAACCC CAGAAGCAAG CAGAGTAGAC ATCAGTCATG
11751 GGTGTGGGAG AGGCAGGAGG GAGAGAGGAA TGGAGGAAGC AAAGAAGGGA
11801 AGGAGGGAGG GAGGGGAGGC TCTAAAACCG TCATCCCTAT TCCAATATCT
11851 GATCTTGAAT TGGCCTCAAC ACCTGTGCAT CCCTGCAGGG GTGGACCCAG
11901 TCCCCAGTTG CTTCCCAGGG AGTACGGGGG TGGGGTGGGG ATTCTCTGGC
11951 TTTCCTCCCT GCCCCTCCTC TGCAGGCTGA TGCTGGGCTT CATGGGCGTC
12001 ACAGCCCTCC TGTCCATGTG GATCAGTAAC ACGGCAACCA CGGCCATGAT
12051 GGTGCCCATC GTGGAGGCCA TATTGCAGCA GATGGAAGCC ACAAGCGCAG
12101 CCACCGAGGC CGGCCTGGAG CTGGTGGACA AGGGCAAGGC CAAGGAGCTG
12151 CCAGGTGAGC CCCTGGCCAG GGCACTGCCA GGCCACAACA GCAGCCTTCC
12201 CCTCCCTCTG CTGGCAAATG CTTTGGCCAC CTCCTTCTCC CTGTCTGCTT
12251 CCCGGAGCCC TCCTTTAAAC ACGCATAGAG AAAAAAAAAT AGAAAATACT
12301 GTTGTCCTAA GTTTTAGGAG GGGATTATTG CACACAACTT AGATCCTTTA
12351 ATAGAGCTTT GAACAAAGTC TCACCCTCAG TTCCCATCAG TTGCAGAAAT
12401 CAGTGTGTTC ACCTGATTAT TCATTTGGGC ATCTTTCGAG CACTTAGGGA
12451 TGCCCCTCAC TCCTTGCTAC TCCTGCTCAT CCTCAAGGAG GCCTTTTCTG
12501 ACCTCCTCGA GCAGCTCAAA TCCTTCCACT CTCTGCTCCC ATAGGTCTGG
12551 GGCTTGGCGT CCCATGCTTG CTTCCCTGCT AGGTGCGAAG CTCAGGGAAG
12601 ACGAGTCAGC ATCTACCTTG CCGTCTGCCG TGTTCCCTTA CCATCCCCAG
12651 CCCAGTGCAG TAGAGTCAGG GTCTGTGGCT GACGGCCTGA TTGCCAGACC
12701 CTGGGCAAGG TCCTGGGGCT TACAGAGAGG AATCGGGCAC ATCCCTGCCA
12751 GCAACTCTTA TGGAGCCCAG TGGGGCAGCT AAATCAGCAG AGCTGGGATT
12801 TCCCAATCCT CAGGTCAGCA GCAGAGTCAG GACCTGGGGC TGGGTGGGCA
12851 GCCCCCATGA CTGGCTCAGC TAACAGCGCT GTGCCCACCA CAGGGAGTCA
12901 AGTGATTTTT GAAGGCCCCA CTCTGGGGCA GCAGGAAGAC CAAGAGCGGA
12951 AGAGGTTGTG TAAGGCCATG ACCCTGTGCA TCTGCTACGC GGCCAGCATC
13001 GGGGGCACCG CCACCCTGAC CGGGACGGGA CCCAACGTGG TGCTCCTGGG
13051 CCAGATGAAC GAGTGAGTCC TTGGTCGCAC CTTCTGGGGA CAACGAAGTG
13101 GGTACCGGGG CTGGAGGGAC CTGCCCACCT CTCTCTGCTC CTCTGCAGAG
13151 TCCTGGAAAG CCTCGGGGCA GCCAGACCTG GCCTGGGAGC CTGGCAGGGG
13201 TGGAAAGATG TGGCCCCATC TAGCCTCTGT GTCCTGGCAC CCCTGTGCCC
13251 ACACAGAAGC CTTAGAGAGG ATAGGGAGCT GATGTCAGGG GAGCTAACGT
13301 CCCAGTCTGC TTTCTGCTAT GATGCAAGAC CCACCACCTC CCCTGGGGTC
13351 AGGGACTCTG GCTCAGAGAG GGAGTGTGGA TTGAACTCTG AGCTAAAGTC
13401 ATGGCAGATG ACAATGTACT TCCAGACGCT GGGTCCTTGG TTGAAACTTG
13451 TAGAAAATAG ACACCTCTAA AAGACTCCCC AGCACTCCCT TTGCTCACTG
13501 CTTTTGGTGG CTAATGGTGA TGGCCCCATG GCATCCGAGG TCTACAGATG
13551 GTATGAAGGG CTGGGGTTGG GTCATTCACT GCTTCACTGC TTCGTTATAG
13601 TCCCCTTGTG AGGTATCAGG TGAACCATGG GATGGTTTGG AACTTTCTAG
13651 CCTTGGCCAC AAAGGGATGC AGGCCATGAG GACCCCAAGA GGGAGAGAAA
13701 CCTGGGCCCT GCCGCGGGGT AGTCATGGTC TGTTGAGGGT GGCAAGATGC
13751 CTGGGGCTTC CAGGCATGTC TGGTACATAA ATGTACTAAT TGAGGTATGT
13801 ACTAATTGCA GTGGGCAGGC ACAAAAATAA GGTGATGCCA TCCTTTGCAG
13851 ACAGGAGCCT GGACAGGGGT GGGGAGGGCA GTGGGCGCAG GAGCTGGGAG
13901 GTGGAAAGGA CAGGTCTGGA GCCTGGCTGG GCAGAAACGT GAGGTTCAAC
13951 AACCCGTTTG TTTTAATTTC GGGAGTGTTT TCTGTAATGA TATCCTTACA
14001 GTTCTCCAGT AACTTTCTTT GGGAAGAGCA GCCCGTCTGG GCTGAGTGGG
14051 GAAAGCTCTG CGCCTGCTTT GACACTCTTG AGCTAAAGGG GGCGCCCCTG
14101 GGGCTAGCAG AGCCCCGGGG ATGGGAGGCG GGGCCTGTGG TGGAAGTGAC
14151 CCTCCTCCAG CCTCCGCTCT GGGAAGCTTT TGAGATTTCC TTTGCTAAGT
14201 GGGGGGACCG TTCTTTGCAG AAACCCACAG AGCGAGATTG CTGAGGTCTC
14251 TGCAGATCCC CAAAGATGTC AGCCAAATTA CATGCATGTG TATAAAAGGT
14301 GTATTTTTCT TTTTTTTCTT TTTGAGACAA GTCTCGCTCT GTCGCCCAGG
14351 CTGGAGTGCA GTGGCGCGAT GTTGGCTCAC TGCAACCTCT GCCTCCTGGG
14401 TTCAAGCGAT TCTCCCGCTT CAGCCTCCCT ATTAGCTGGG ATTACAGGCG
14451 CCCGCCACCA TGCCTGGTTA ATTTTTGTAT TTTTAGTGGA GACGGGGTTT
14501 CACCATGTTG GCCAGGCCAG TCTTAAGCTC CTGACCTTGT GCCCCACCTG
14551 CCTCGGCCTC CCAGAGTCCT GGAATTACAG GCGTGAGCCC CTGCGCCCGG
14601 CCACAAAGTT GTATTTTTCT GGAGGGATGG GCCATAACTT CCATGAGACT
14651 CTTAGCAAGG CCTGGACACA CAGAAGAGTC AGTGGGTCAT TTCTCGGCCT
14701 TGTCTTGTGC TGTGGCCATG TTCTGAGGCT CCCACTCGAT TAGGGGACAA
14751 TGCTTGGCAA TGGACTTGGT GGCTAGACCT CAGGAGGATG TGGCCTCCAC
14801 ACAGGCGCGC CTCTCAGGGC CCAGCTGCTG CTCCGTCCCC ACGCACAGGG
14851 CCAGGCTGGC TCCCACAGCT CAGCATCTGA GGTGGGGGCC GGTGTCTTCT
14901 TGTAGGTTGT TTCCTGACAG CAAGGACCTC GTGAACTTTG CTTCCTGGTT
14951 TGCATTTGCC TTTCCCAACA TGCTGGTGAT GCTGCTGTTC GCCTGGCTGT
15001 GGCTCCAGTT TGTTTACATG AGATTCAAGT AAGTTTGAGC TGCTCACAGC
15051 CTAATTATGC CTAATTATGC CTCAAAGCTG CAGAAGAGCC CTCAGACTCA
15101 ATAGGCAGGT TTACAAAGTC CTTCGTGTCT GGCCCTGATC TTTCTCCAGC

FIGURE 3D

```
15151 CCTGTCTCCT GCTAGTCTGC CCTCCTGTTC CTTCGAACCC AGGCTGCTCA
15201 CTGAGCTTTG TGCACACGTG GTCCCCTTTC CCTGGAATGC CATTCTCTAC
15251 CTTCCCACCT CCTCAGCCTT CAAGGCTAGT TCAAATGCTG CTTCCCTGAC
15301 TTTTCCCCAC CCCCATTCCA TCTCTGAGCG GCCCCTGGGC ATATCACAGG
15351 CCTGTCCTTT AGTATCTGCA TTTGGCTTCC GGTGACTTTG AATTCCTCCA
15401 GAACCACTCT GATGCTGGGC ACCCCGCACA GCTCCCAGCA CAGGGAGGAA
15451 GAGCAGGCAG GTTAAAGCAA TTAAAGATAA GCTGGTCCCC ACGTGCCAGT
15501 TCGACATTGC TGGACAAGCT TCCTCTTTGC CGTGTGGGTC CATCAGGCCA
15551 GGTCACCGCA AACCTGTGAC TTAGCTCTGA GCTGAGCGCA TACGCTCTGT
15601 GCCTCAATGC ACGGGGAGTT TAAGTCGAGT AAAACCAGCA GTGATTATGA
15651 CCAAATCCAT CCAAACCCAG ACATTTACTG AATACCTCTG GTGTTCCCAG
15701 CAGTGTACAG GTCCTAGAAA GTTTACCTTC CTGTTCCTAG CACACAGGCA
15751 AGTTCATCAG GGGTCACCTT TGATGGCAGC CAGACTTTGG ACAGAAACCA
15801 TGACCTGTGG CTGACAAATA GCTAAAAAAA AGTTATTGTT TTTCTAAAAC
15851 ACACAAATTT ATCTGTGGTG CAAAGGTGAT CAGGCCACAC CAGGATAGAA
15901 AGTACTCAGC TCTGAGTTAA GTGCCTGTGC TCTGTGCCTC CATCCACAGG
15951 AAGTTCGAGC CAAGTCAAAC CAGGGGAATT TGTGACCAGA GGGAAGAGAC
16001 TGCAGAGCTC AGAGGCAAAA GTGCCCACGG AAACCTGTGA TTTTGTGGGG
16051 AAAATAGGGA ATTTTCCTAA GTTTTCTTCT GAAGGAGGAA CTGTTTTGAA
16101 AACTCCCATT AAAAAGTTGC TATACAGGCC GGGCGCGATG GCTCACACCT
16151 GTAATCCCAA CATTTTTGGA GGCCGAGGTG GGCAGATCGC CTGAGGTCAG
16201 GAGTTTGTAA CCAGCCTGGC CAACATGGTG AAACCCCGTC TCTACTAAAA
16251 ATACAAAAAT TAGCCGGGCG TGGTAGCCCA CGCCTGAAAT CCCAGCACTT
16301 TGGGAGGCCA AGGAGGGCGG ATCGCCTGAG GTCAGGAGCT CGAGACCAGC
16351 CTGGCCAACA TGGTGAAACC CCATCTCTAC TAAAAATACA AAAGTTAGCT
16401 GGGCATGGTG GCACATGCCT GTAACCCCAG CTACTTGGGA GGCTGAGGCA
16451 GGAGAATTGC TTGAAGCCGG GAGGTAGAGG TTGCAGTAAG CCAAGATCAT
16501 GCCACTGCAC TCCAGCCTGG GCGACAGAGC AAGACTCTGT CTCAAAACAA
16551 AAAAAAAAGT TGCTATACAT ATTCAAAACA ATCATAATAA TGATAGTAAG
16601 AATGACAATA TTAATGATCA TTGCCCAAAC CCCACTCTGT CCTGCCCATG
16651 GACGGGGCAG GGGAAACTGT TTGCATGGCT GCCTGGCCAC CCAGCCTGGC
16701 TTTGACAGTA GCTCTCTTTG CCCTGCCTCT TGAATCTGCA CCAGGGCCAA
16751 AGTCCTGTTC ATTTGTTCAC ATCCGTCGAA CAGGTCTCTC AGGAGATGGT
16801 CCTGAACCTG CTGCAGGTGA GCATCTGTGT CTCCTCATGG GGCAACAGGA
16851 ATAATAATGA CCAACATTTA TTGAGTGCTC ATCATGTGCC AGACATGATT
16901 TCGAGCGCTC TTTTCCTTTC TTTATTTTAT TTTATTTTAT TTTATTTATT
16951 TATTTATTTA TTTATTTATT TATTTATTTA TTTATTTTTG AGACAGTGTC
17001 TTGCTCTGTC ACCCATGCTG GAGTGCAGTG GTATGATCTC GGCTCGCTGC
17051 AACCTCCACC ACCTGGGTTC AAGCAATTCC CCCTGCCTCA GCCTCCCAAG
17101 TAGCTGGAAT TACAGGCACC CACCACCACC ATGCCTGGCT AATTTTTGTA
17151 TTTTTTAGTA GAGATGGGGT TTTGCCATGT TGGCCAGGCT GGTCTTGAAC
17201 TCCTAACCTC CGGTGATCCG CCCTCCTTGG CCTCCCAAAG TGCTGGCGTT
17251 ACAGATGTGA GCCACCTCGC CTGGCCCAAG CACTCTTAAA CTTAATTAAT
17301 TTTCACAACA ACCTGCGAGG TCAGCACTAT TATTATTATT CCCAATTTAC
17351 AGACAAGGCA ACTGAGGCAT GGAGAGGTGA TGTGGTCAAC ACAGAGCTTT
17401 GTAACAGGGA AGTAGGGGGA CTGAGACTTG AACCCAGGCC CTTTGGCTCC
17451 CACTGCATGG CATCCCCTCT TGGGGAGGCT GAGGGTTGCT GTCCTTAGTT
17501 GCCTCCAGAC CTAAGCATGA CCAGGTGTCA GAAACACTAG TTGGGGCCGG
17551 GGCTGCCCTA GAACCCCAAG GCCTACTGAG AAAGAGGAGG GAGATAGCAT
17601 GGCGCCGAGG CCGCAAGGGC ACCATCAGCT TCTTGTCTGG CCAGAGGCAG
17651 ATGTCAGGCC CCTGGAGACT CACAGCCAGA ACCTGAAGCT GAGTCCACCC
17701 AGCCTGGCAC GGCCTTCATC AGCTTTTGTT GACTGGCGGG GGAGCCTGAG
17751 AGTGTCTGCA GCAGGGGGCT TCTGAGCATG CTCGTGGTGG GGTGCGTGGC
17801 TGCAGTCCAG TCCCACCCCT TCCCCTTCCC GACGGGCCAC TCTAGTTTGG
17851 ACGCATGCAG TGTGGCTGGC CGGGGTAGCT CACGGCAGCT TTGTTTTGGC
17901 TCCAGATCTG GAAGGTAGAG GACAGCTTTT ACATTCGGTT TGAGTGGTGG
17951 GAACAGTGCT CTGGCCCAGG CCACGTCCTG CCACAAACTA AGACCTGGTG
18001 GTCCCTGCCT GCCTTTGTGG CCTCATGGAC CTCCCCACCT GAGGCCAGGG
18051 AGCACCTGTC TCAGCGGCAG GAGGCAGCTC CACTGTCAGC TGTTGCTCTC
18101 ACTAGAGTTC CTCATCTGAA CGATCCTGGA GAACGAGGTT AAGTTCTTGG
18151 CCTCTAGCCT AATCCAGAAC AACTATCTTG CTGAAGAGCC TAGTGCAGCC
18201 TCCTAGGCTA TATCTAGCCA AAGGGGCCAG ACCCCACCCC AGGACCACCA
18251 AGAACTACAT GGGATATTAT TACTGGTTAT ACCTAACTGT CCCAACCAGG
18301 CTTACCTCCT GTAATAGCCA TGAGGGTTCT TTGGGACCCC TGCCAGGGCA
18351 GAGGCATGCA AAGCTCAAGA ATCTCTCCCC TCTTGTTGGC TCTGCAACAT
18401 ATTCAGTCCA AGTTCACCAT GGTGCATCAT GGTGAAGGCT GTTCTGCTGC
18451 AGGAGGACTC TGTGGTCCCC ACCCCTGACC CTGACCTAGG CCCCTCACAG
18501 GCCAACTGGA TCCATTTACT TGCATCTCAT GCCAGCCTGG TCATCACCAG
18551 ATGAAATTAA CCCAGAGATG AGAGCAAAGC TGCTCAGCAC GAGAGACTCT
18601 GAAGGCTTGG CGGTACCACT GTGGGGCACT GGCATTGGAA GACTGCATAC
18651 TCCATGCAGC CCCAGAGTCT GCAGCTACTG TGGTGTTGGG GATGAGCTGC
18701 CAGCACCAAA TGCAGGCTCT GGCTCCTGGG CCACTAGTAA TACCAAGGTC
18751 ACCCCTTATG CTGGAAACCT GAAGCCCCTG GCTGAGCCCC AGGGTCTCTA
18801 GGACGACAGT TGGCAGCAGA GAGGTGCTTG GTAGAGCACA AACTTTACTA
18851 AGCCAAGGGT GTGGCAGCAG AGAGGCCCTG TCTTACACCA GCAGAGCCAT
18901 CCCTGTGCCG GATGTCTAGA GAGTGTCCCT AGCGGGTGAC CCTCAGGACA
```

FIGURE 3E

```
18951 CACGGCCTTG CCCAGCAGGG AGATCCTAGC CAGCCGTGTA GACCTGAGGT
19001 CCCATCAGTG TTGCCTCCTT TTCTGACCCC TGAGCACCCC AGAAAGCTGT
19051 GACCTGATGT CCTGGTGTCC CCATGTTCCA GGCCAAGCCA CCATCACACC
19101 AACACTTGGC CCTCACACTC TCCAAGGCTG TTCACATCCA GCACTGGCTT
19151 CAGGAATGAG CTCCTATTCC ATCAACCCCT TCCCTCCTAT GATTATGTCT
19201 CATGGCCCCC GGGAAGGGCT CTCACGAGGG AGGGCTCTCC AGGACAATAC
19251 TCTTGGCCTT GCCCACCCCT TCAAACCAAC AGTGGCTGGA ACTGGAATGT
19301 GTGAATGGAA TATTCAGCAT ACCTTGAGGC CTTAGTCCTA TGCACAGTGG
19351 CCCCAGTTAT CCCCCCTCCA CAGCTGAGCT CCCCTTTACA CCTCCTCCAA
19401 GAACCTCCTC TCCTCCCTGC CTCCTCATGC CAACGCCACC TTAGGGGAGG
19451 CCCTGCAGGA CACCCTGGAC AATGGACACT GGTCCCAAGG GGGCCCATCC
19501 AGGATGGGGG TGCCATCCTG GGCTGTCTTC CTTCTTGCCC TAGCCATGCT
19551 TGCTGCTAAC CCCAGGGTCT CCTGGATCCC TAATCCTGCA CCTCCAACTC
19601 CAGGGAACAC AAGGACCCAT TCTGCCCCTG ACTAGCCCTG TCTGCCAGGG
19651 TTCATACTCA CTCCCTGCAT CTCCCTGAGC CACCTTGGTG ATGGGGGTTG
19701 GCATCCCAAC ACCATCGAAG GCAGCTCCAG GCTGAGGTGG AAGGAGGAAG
19751 ACTTGGGAAG CATGTGAGGG AGCCCTGTTC CCACCTTGCG CAGGCTCCGA
19801 AGCTCCTTAT GGCCTTCCCC CAGGTGACCC TGGAGCAGCC AGTCTCCAGG
19851 TGTCTGGGCA CCTGCCGAGA CCCTCTAGCC TCTCTACAGA GACTTTTTCC
19901 CTAGTACATT CTGGGATGGA AGAACAGGAG AGGGAAAGAG GCAGGAAGGG
19951 CCTTTCTCCA GGCCCCATAG CAGGCGAGGA CAGCATTATG TGTCTTTTTG
20001 CTACATTCTG CTGTAGAACA TTTAGGCTCC ATCTGACCAG CACCTGAGCC
20051 AACCAGTCTG CCCTGCCCTT CTCTCATCTT TGCATTCTCC AGTTTTAAAA
20101 AGTCCTGGGG CTGCGGGCTA GAGAGCAAGA AAAACGAGAA GGCTGCCCTC
20151 AAGGTGCTGC AGGAGGAGTA CCGGAAGCTG GGGCCCTTGT CCTTCGCGGA
20201 GATCAACGTG CTGATCTGCT TCTTCCTGCT GGTCATCCTG TGGTTCTCCC
20251 GAGACCCCGG CTTCATGCCC GGCTGGCTGA CTGTTGCCTG GGTGGAGGGT
20301 GAGACAAAGT AAGTCTTGGA TTCAATAGAA ATCGCTGGCT TAGGGCCAGG
20351 CGCGTTGGCT CACACATGTA ATCCCAGCAC TTTGGGAGGC TGAGGTGGGT
20401 GGGTCACTTG AGGTCAGGAG ATCGAGACCA TCCTGGCCAA CATGGTGAAA
20451 CCCTGTCTCT ACTAAAAATA CAGAAAATTA GCGAGGCATG GTGGCACATG
20501 CCTGTAGTCC CAGCTACTTG GGAGACTGAG GCAGGAGAAT CACTTGAACC
20551 CAGGAGGCAG AGGTTGCAGT GAGCCCAGAT CGTGCCACTG CACTCCAGCC
20601 TGGGCAACAG AGAGAGACTC CGTCTCAAAA AAAGAGAAAG AAAGACACCA
20651 CTGGCTTAGT GCACTAGTGC CTAAATGCTG CTGGTCTCGG CTACAGGTGG
20701 CAAGAGGAAT GTGGGCCAGG CACTCATGCT TGGTCAAGAC TTTTCCTCTT
20751 TTGGGAGCTG GGTTTCAGAG AGCACTCTGT TGGTTTCATG ACTCATTTTT
20801 GTTTTCTGAC CAAGCTCCAC AATAAGACCC TAATGTGTTC CTGTGGTATC
20851 CTCTCCTCCC TGAGTAGGCT GAGCAGAAAA TCCTTGGCCA GGCAGGGTGG
20901 CCAGAGCTGT GATGAGAGAG ATTTCTTGGG CTAGGAGTAG GGTTCCCAGA
20951 GCTCTAGTTT CCAAATCTCT GCTCTGCCAT CTTCCCTTTC TCATCTTCAC
21001 ATCTGGTCAA ATCCCTCCAA AGGCACACAT CTAGGGAGCT TCATAGACAG
21051 AGACTTGGCA AAGGGGGTAC ATGTAGTTTC TCTCCTGGCT AAGACGTTGT
21101 CAGAATGGAA GAAAGGATGA GAAACATGTA CATCCTAGAA AAGGCAGAAG
21151 ATGTGGGCAG GGAGATGCTG GTATGATGGC CATTTCGTTT TGAAGGTCGG
21201 CTTAGGTCAG CACCAAAGTC TTCATGGTCA CCCTGGTGAA CCCAGACAGA
21251 ATTCTAGAGA ACCTGGTCAA GAAGAGGTCC TGAAATACAC TTATGGAGAA
21301 TGCACGCTGA GAGGGGGAAG TAAACTGCTT AGGATCACCC AAAGTTGGTG
21351 GTCAAGAGTG TGGGCATCTT GATTTCTAGC CAGGATTCAG TCTCCCATAC
21401 CACTCTTATT TTTTTATTTT TTTGAGACAG AGTCTCACAC TGTCACCCAG
21451 GCTGGAGTGC AATGGCATGA TCTCAACTCA CTGCAACCTC CACCTCCCAG
21501 CAATTCTCCT GCCTCAGCCT GCCGAGTAGC TGGGATTACA GGCGCCCGCC
21551 AGCATGTCTG GCTAATTTTT TGTATTTTTA GTAGAGACGG GGTTTCACTA
21601 TGTTGGCCAG GCTGGTCTTG AACTCCTGAC CTCGTGATCC GCCCGCCTCA
21651 GCCTCCCAAA GTGCTGGGAT TACAAGTGTG AGCCACTGCA CCTGGCCACC
21701 ACTCTTGACC TTGACTTTTA AGGCTGTGAG CCTGTTTCTT TGCATAGAAG
21751 CATTTGGACA CAGAACTGCC GGAGTTGTGA TGGGTTTGTT GAGTGACTGT
21801 CTCTGTCGCA GATGAGCTGT GCTTTTCCCC ACCTAGGTAT GTCTCCGATG
21851 CCACTGTGGC CATCTTTGTG GCCACCCTGC TATTCATTGT GCCTTCACAG
21901 AAGCCCAAGT TTAACTTCCG CAGCCAGACT GAGGAAGGTA AGTCTCCTGT
21951 TCTGATCGCC CAGTCATCAG GACTGGAGCC CTGGAACCAA AGGGTCACTA
22001 TGGGATGCCT TGGGCCCTAG AGGGAGAAAA TCCCATCATA TCCAAGAGGA
22051 TTGGCTACAA AAGCCTGGGA AACAGTGGCT TTCAAGCCAC CGGTGGTATT
22101 ATTTAGTGCA AAATATCTTT TTTGCTTTTT AACATTTGAA TTTAACATTT
22151 GAAATTTTAT TTATATTACA ACAGGAACAG AAAATGTTTC AAATTTTCCA
22201 TAACACTGAT TTCCATTCAG CACAATTTTT TGTTTTCTCT CTTCCTCCCA
22251 GTCTTTGCTA ATATGCCTGT ATATTACATT ATAATCAACA CACACAGTTT
22301 GAATCCTATT TGTTTGTTGT TTCTTCTACC ACTTTTGATT GATATTACAC
22351 TATAAACATT TCCCACTATT GCTACAGTCT TCAAATATAT TTTCTCTAAT
22401 AATGGCATTA TATTGCGTTG AGGGGTTGTA ATCATTCTCC TGTTATAGAA
22451 CATTTTGGCT GTTTTGAATT TTTTATTTTC ATAAATTAAT GTTTTCTTGC
22501 ATATAGCTTT TCCTTTGAGG GTATTTTTTC TTTAGGATAA ACTTCTAGGA
22551 GTAATATTGC TGGGGTGATA GAATACAAAG TCTTAATGGC CCTTAAAATG
22601 TATGGCCAAA TTGCTTTTCA AAAAGGTCAT ACCAATTTAC GATGCTATTG
22651 GCAGTGTGTG TAATAGTTTG ATCATATCCT CACCAGCAAT GTATATATTA
22701 TTGTAAACTT TAGCTAATTT ATAAGTAGGA GATGGTACCT CATTGTCCTT
```

FIGURE 3F

```
22751 ATTAGCTTTA TTCCCCCTTG ATTAGATTTC TTTTGTCTTC TAATGCTGCT
22801 CTGGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTTTCTT
22851 TTTCTTCCCA GAAAGGAAAA CTCCATTTTA TCCCCCTCCC CTGCTGGATT
22901 GGAAGGTAAC CCAGGAGAAA GTGCCCTGGG GCATCGTGCT GCTACTAGGG
22951 GGCGGATTTG CTCTGGCTAA AGGATCCGAG GTAACTTCTC CAGCCACAGG
23001 CTGCCCAGAG CCCTCTTCTT CGTCAAGAGG GTGGCGTTTC TCCACCCTTC
23051 CATCCCTGGG CTTGTGTGTT TCTGTGCCTG CATCCTTCGT ATAACCGCAC
23101 ATTCCTTGAG GACATGGACT CTGTCTTGTC ATCTAGGAAC TCTACACCAC
23151 ACACAGGGCC TGGAAGACAG AAAGTAACCT TTGAGCGATT GCAGGAATGA
23201 GTGAATGAGT GACCGTGGTT AGCCAAGAGA GGCAGAGGAC ACTGTCAGTT
23251 ACCCTCTGGG GCTTGATCAC AATAATCTCT GCTTTGATTT GTCTGAGGGA
23301 AATCTTTCTT TCCAATCCTT GTCAATATTG TTTGCTACTA CTTTTGGTCC
23351 TTCTACTGGC TACTTAACAT GGTAGCTACT TCAAAATTTT TCTTTAGCTA
23401 AGTATGTAGC AGCGTAGGAG GTGAGGAACA TGTTGGAAAA CACACAAAAA
23451 TATAACTTTC TTTACCTCCT TCTTTCCCTC CTGGGGAAGA AATGAGCCAG
23501 AGGGAGGGAT GAGCTAGCTT GCTGCTGCTG TCCTCCAACC AACCATCTAC
23551 CTACCCAAGT ATCCAGGAGT GTAATAGACA GACTTGGTCT AGTTATTGCT
23601 GTTTCTTCAA TATCTAGGAC ACAGCCTGGT GCCTAGTGGG TGCTAAGTTT
23651 TTGCGGAGGT GAACAAATCC ATCCATCTAG TCACCTCTCC ATCCATCATC
23701 CATGCATCTA TTCATGTCTT CATCCATCCA TCCGTCCGTC CATCCGTCCA
23751 TCCATCCATC CATCCATCCA TCCATCCATC CATCCATCCA TCTGTCCATC
23801 CATCCACCCA TCTATCTATC CACCCATCTA TCTATCCATC CATCCATCCA
23851 CCCATCTATC CATCCATCCA TCCATCCATC CATCCATCCA TCCATCCATC
23901 TATCCATCCA TCCATCCAAC ATTCCTATTA TGTCCCTAGT GTTATGCCAG
23951 GCACAGAGAT TACAGAGGAG ATTGAGATAC GGTCCCTGTT CGTGGCAGAC
24001 TTCACAGACT AGGGAGGGGC ACATATATGA AAGGGCATTT CAGGAAGTAG
24051 CACACGAGCA AGGGAAAAAT GTGAGGTATT TAGCTGAGGA GAAGTAGAAG
24101 ATGAGGCTGG TAAGGCTACC AGAAGCCACT TCTCTGAGGG CCCCAAGATA
24151 GAGGGGTGTG GACTTGATCG TGAATGCAGT AGACAGCCAC TGAAGGACTG
24201 AGGCCAGGGG GTGAGTTGGT CAGATCTGCA CATGAGGAAA TCACTCTGAT
24251 GTCTGGAGTG GGGGCCTGGG CTGGGCAGGG CTTGGAGGAG AACTAGCTGA
24301 GACTCTGCAG CCTTCCATCT CACTCAGGCT CAGAACTTTG GACTCTGTGG
24351 ACATTCTCTC CTCCTTTGGC CCCCAGCTCA GCACAGTCTC CAGCTTTACT
24401 TCGGACTCAG ACTATTCCTG CTCAGCCTTC GTTGCTGACT TCTCTGTTCT
24451 CCCTGAAACA GGAGTGTCTG CCCAGGCTCT GTCCTTGGCC TCTCCTCTTT
24501 TTACACTTCA TTCTCTCCCT GGACAATCTC TTCTCAGCCC AAAGCCCTAA
24551 ATCTAAACCT TCAATTTCTG GTTGAAATCA TTCTCCTGAG CTTCCAAAAC
24601 TGTGGAGCAC TGAAGAGGAG GAGATGGATG TGAGACATTT GGGTGACTTG
24651 GTGACTGACT GGGTATAAGG AAGGAGGGGA ACAGAGACCG GCAGCATGAC
24701 TCCCAGCCTG CTGGGCTGGA TGGCTGGTGG ATGGTGAGTC CATTCACCAA
24751 ACTGGGAGGC CCAGAGAGAG AAGCAGATTC TGGGCTATGG AGGATGAATG
24801 CAGGGTGGAG CATGTTGAGT CTGTTGTGCT CTTGGGACAT CTGGATGGAC
24851 ATTTCCAGAA GGCATATGGG TATGTAAATC CACATAGTAG GCCAGCTGGC
24901 TGGAAATACA GATTTAGGAG ACAGCAGAGT GAGGACGGGG ATGAAAATGG
24951 TGGGAATGGA TGAGGTCACC TATGAGGTGT AGAGAGAGAG GGTCGGGAGG
25001 GGAGGATGGG CCAAGCTTGC CCTGGCCCTG AAGGAACTGC AAGCTGGGAG
25051 CGCTGAGATG ACTGCCCTCC TGGTGCTTCC CAGGCCTCGG GGCTGTCCGT
25101 GTGGATGGGG AAGCAGATGG AGCCCTTGCA CGCAGTGCCC CCGGCAGCCA
25151 TCACCTTGAT CTTGTCCTTG CTCGTTGCCG TGTTCACTGA GTGCACAAGC
25201 AACGTGGCCA CCACCACCTT GTTCCTGCCC ATCTTTGCCT CCATGGTAAG
25251 TAACCTGACA GTGGGGAGGA GCCCTTCCAT TTCACAGGAA CACATGGCCA
25301 TATTGTGGGT CCCTGACGAG GCAGCAATGT CCAGGCCAGA CTCAGACCAG
25351 GCTTTGGAGA CCCAGGTCTG ACTGTGACGT GGATTTGTGG ACCCTGGATG
25401 CCTCTGCCCC TGAGGCCTCC ACTGCTTTGC CACTCCTCTT TGTACCCCTC
25451 CTGCTGACCA AAGCACCAAC CATGGACCAA GTGCTCAAAT TTATTTTATA
25501 AATCTAATTG GATTATTTTT CAAGCTGGGG AGACAGGACT TGGGCTAAGG
25551 AGGAGCAGGC CAGTGCCGTG GTCTCTGAGC ATGTAGCACA GGTGTGCAGG
25601 AGGACTGCAG ACTGGGAGCA CCACTGGCTG GAAACCCCAG GAAGAGGCCT
25651 TGGAGGAGTG GGGACTTGGG AGTAGGTAGG AAGGGAGAGA GAATTCTGGG
25701 AAGATGGAGC AGCACAAGGA AAGGCAATGG TGCACATGAC TGAGGACTCC
25751 TGGAAGCCTG GCTTGGTGAG CACAGGGATA AGGGATCCTG GGGAGTGGAG
25801 AGAGGTAGCT GTCGGTTGTG GGAAAAGCTG CTGAGTGCCA GGCTAAGGCA
25851 TTCTGTTCTA TGGACTAGCA TGTTTTTTAG TTGGGAGTTA GAAGAAAGCA
25901 GAGCTTATAG GAAAATCAGT GGCTATGGTT TTTTTTTTTT TTTTTTTTTT
25951 TTTTTTTTTT ATGCATTTCC TTCTGTCATC CATTGCAAAG ACGTACCAGC
26001 TTCAGGGTAG TATGGAAAGA TCCCTGGTCT CGCAGTCAGA AGACCCGAGT
26051 TCAAGATGTG GGATCTCTGA ACATGGCCCT TCAGTTCTTT CTTCCGAGAG
26101 CTGTGCTGAT GGCCAAGTAA GATGAGGGCT ATGAAAAGCC TCTGTAGACT
26151 GCAAAATGAG CATGGGAGAG GCTGTCATTA TTCTGGAATT GGGAGACAGA
26201 TTTACAGAGG GCCTGAACAC AGGATTGAAG GTGGTGAATT TCCATTCGGC
26251 TGCCTGGGCG TCTGCATGTA TAAAAAGCAA ACCTAAGTGG TTTTTTTCTC
26301 CTCCAAGTGA AGATGAAAGT GTTAAAAATA GCAAGGAGGT GAAAGTGTTC
26351 AAAATAGCAA AGTGGCCTGT CTCCTCTTCT CCTAAGCAGA CTGTCCAAAC
26401 AGACGCCCAG TAGAAGGAGC ACCTTTTGAT ACTGGGCACG TGGTGGTGAT
26451 GCCTCCTCTC TCCTAGCACA GGCCGTGGCT TTGTCATCTC CAGCCCTAAC
26501 TGGGAGCACC GAGGGTTCCA ACCAGGCAAA TGCAGGCCCT AACGGGCTCT
```

FIGURE 3G

```
26551 TTGAAAACGG GCTTTTCTAG AACCAGGAAC CTCAAGTAAA AACTCCCCCA
26601 GCTACCTCTA AGGCCCATCA CACTCCTGTC TCACGCCCAC CTATGAGAAA
26651 GGAAAAGGTG ATGGTCATTG AGCTGGGCTG CAGAGGAGTG TGAGGTGCAG
26701 ACACCATGAG GTACCCACAG CCAGGAAAAC GAGGATGGTC GGGGAGACGC
26751 GCCAGCGAAG AGCTGGGCCC CTGCGTGGGA CCCCTCAGTG GTTCCCAGGG
26801 GGCGTGGGAC TTGCGCAGTC CTTTCAGAGG GCTGTTTACC AACAGGAACC
26851 GTAACATTAA ACCTGCTCAG ACCCCTTGAC TCAGCAATTT CATGTCTGGG
26901 AATATATCTT AGGAAAATAA TCAGAGATGC CTACCAACAT ATGTGATGAT
26951 GATGTATGAC AGAATTATTA TACAAATATA TCCATAGTAA CAGGGGGTTT
27001 GCTGAAATAA ATTATCATAT ATTCATATAA TATGACATTA TCAGGCCATT
27051 AAAAATCACA GTTTCAAAGA GTAATAAAAT GGGAACATGC TCATAGTATA
27101 GTTTTTTAAA ATTGCAGATG GTATATGGCT AAAAATGTCT AATAATGCAA
27151 AGATGTATAC AGACCTTAAT CCTCTAGCCT CCTCCCTAGA GATGACCTCT
27201 GTTAATTTCT CAAATATTTT TCTGGATATT TTACACACTC ACACACTTTT
27251 TTTGAGACAG AGTTTCACTC TTGTCACCCA GGCTGGAGTG CAATGGTGTG
27301 ATCTTGGCTC ACTGCAACCT CCACCTCCCG GGTTCAAGAG ATTCTCCTGC
27351 CTCAGCCTCC CGAGTAGCTG GGATTACAGG TGCCTGCCAC CTTGCCTGGC
27401 TAATTTTTTG TATTTTTAGT AGAGACGGGG TTTCACCACA TTGGTCAGGC
27451 TGGTCTCAAA CTCCTGACCT CAGGTGATCC GCCTGCCTTG GCCTCCCAAA
27501 GTGCTGGGAT TACAGGCGTG AGCCACTGCG CCCGGCCATT CATCTTAATT
27551 TTTAAAAAAT CTAACCATGA AGCCTTGGTT ATCTTGGAGA GCTTTCCTGA
27601 TTAGCACAAA AAGAAAAAAA AATCCAATTC TTTACAGCTG CATACTATTC
27651 CATTATTTGT ATGTGTCATA TTTTATTTAA CCATCCTGCT ATTAGTGACC
27701 ATTGAGTTGG CTTCCTGTGT TTTGCCGTTA CATGGTTGCA ACAAACATGT
27751 TTGCATGTGT CTGCCCTCAT GTGCATGATA CATGATTGAT TTGATAGATT
27801 TTAGGAATTA CATCATTCAT TCATACACTC AGCAAATATT TAATGAGTGC
27851 CTACTCTCTG ATAGGTGCTG TTGGATGTGG CTAAATTTTA AAGTGTAGAA
27901 TTTAAAAGGT GGCTACCAAA TTCCATGTGC AAAATGACCC CACGCATGTA
27951 TAAAAACACA CACATCCACA GATTTATATG CGGGAGAGAA GATGTGGTCC
28001 CTGGCCTCTA GGCTCTCTCA GTCTGTGGCA AGACAGACAG ACATGTGCAC
28051 GCGGCACTGT AAGGTTGAGC ACAGTCTAAG TACTCAGCAT GGTCTCTGGC
28101 ACATAGTAGG TGCCCAAGAA ATACATGTCG AATGAATTGA GGGGGTAAGG
28151 CCTTCTAGGG CAGGTGGCCT CTGACCTCAG CCTTCAGTGT TCCGTAGGTG
28201 GAATTATCTG CCAGAGACGT GGCAAAAGGG AGAGGAACCA AGACTGAGGC
28251 ACAGAGGTTC AAACGTACCC GGCACATTCA GAGAATCCTT TTCAGAATCA
28301 CGTCCCCAAG AGCTTCTGTG TTCTGTACGG TGATGTTGCA GTGCTGTTTT
28351 TCCGCAGTCT CGCTCCATCG GCCTCAATCC GCTGTACATC ATGCTGCCCT
28401 GTACCCTGAG TGCCTCCTTT GCCTTCATGT TGCCTGTGGC CACCCCTCCA
28451 AATGCCATCG TGTTCACCTA TGGGCACCTC AAGGTTGCTG ACATGGTAAC
28501 ACAGCTGTTT TTATTTACTC CCGTCGGACT ATAACGCTGT TGTCATAAGG
28551 GATGCCCCAT TTATGAATGA CAGAGTTTCA AAACGATGTC ATGTGACTTG
28601 GGAATGCCAC GGAACATCCA GACCTGTAGC CATTGTTGAC ATTTATAATG
28651 CAGCTTTTCT TCTTTTTCTG AGATGATCTC AAGCCTCACA CACTGTTCTT
28701 TCTCTGAGGT GGGTTATAGA CTCTCCCACC TGGAGAAGCC TGTGCAGGCA
28751 CCAGGGGAGT CCTTGGAAGG GGTGAAGGTG GGGCTGAGGG ACTCATATGG
28801 CCAAGGATGA ACTTGACAAA TTAGCAAGAA CCATGAAGAT AGGCAGGGCA
28851 GGCTTAGGCA GCAGGGGGAT GCTAATGACA GTCACAGAGA TTTGTAGGGG
28901 TGCCTGAAGA GGTAGAAGCA GGGAGAGGGA GAGAGAGAGC ACTGCCTGGG
28951 AGTAGATGAT GCCTTGGAAA CAAATGTAGT CAGAGGAAGA ACTCTTCATT
29001 AGCTCTGTCA CCTTTGCTGG GAGAAGGGCA GCTTTGCAGC TCTGGGCTGG
29051 GAAAGAGGCA AGTGTTTGAG CCCAAGAGGC CAGAAATGTA CCTGGGACCA
29101 ATCGGGTGTT CGTTATCTCA GAGCCTCTGC TGGGTATCTC AGGGACTCCA
29151 TGAGCATTTT CAAAAAAAAA GGTGGGTCCC AGAAACCATG GACTGCAAAC
29201 TTGACTCCAA TCCCCCAGTA AAATATCTAC AACAGGGTAG TGAAGCGATG
29251 GTTAGTGACC ATGAGGGAAG CTTGCAGAGC AGGCATCAGA AAGAGCCTGA
29301 GGAGGTCCAC AGGGAAGCTG GCACGTCCTT GTAGGATAGT TAAGGCACTG
29351 GGGTGAGCAA TGAACCTGGA CTCACGGAAC ACTGGGCTCT GTGACCGTTT
29401 CCCTGAATGG CCTAAGCTGT TGCCTCCTGT CACTTCTCTG AGGTCATTTT
29451 CCAAATGCGC ACGGGCATAG AGAACCCATC CACTCTGCCT ACTTCCCAGG
29501 GATGCCTTGA GCACTGAGGA TACCTGGGGG ACATGAAGTC GCACTGTCCT
29551 GGGGGTCGGG ACACCCCAGC CAGGGACAGA GCATGGCACA GGGACATCGA
29601 GGCCCAGTGA GCCGACCCTT TGTCCTCCTC TCTGAGAGCA CTAGTCCCCA
29651 GCAGGCCTCA GGGTGCTGAC TCTGTCTCTT TTCCAGGTGA AAACAGGAGT
29701 CATAATGAAC ATAATTGGAG TCTTCTGTGT GTTTTTGGCT GTCAACACCT
29751 GGGGACGGGC CATATTTGAC TTGGATCATT TCCCTGACTG GGCTAATGTG
29801 ACACATATTG AGACTTAGGA AGAGCCACAA GACCACACAC ACAGCCCTTA
29851 CCCTCCTCAG GACTACCGAA CCTTCTGGCA CACCTTGTAC AGAGTTTTGG
29901 GGTTCACACC CCAAAATGAC CCAACGATGT CCACACACCA CCAAAACCCA
29951 GCCAATGGGC CACCTCTTCC TCCAAGCCCA GATGCAGAGA TGGTCATGGG
30001 CAGCTGGAGG GTAGGCTCAG AAATGAAGGG AACCCCTCAG TGGGCTGCTG
30051 GACCCATCTT TCCCAAGCCT TGCCATTATC TCTGTGAGGG AGGCCAGGTA
30101 GCCGAGGGAT CAGGATGCAG GCTGCTGTAC CCGCTCTGCC TCAAGCATCC
30151 CCCACACAGG GCTCTGGTTT TCACTCGCTT CGTCCTAGAT AGTTTAAATG
30201 GGAATCGGAT CCCCTGGTTG AGAGCTAAGA CAACCACCTA CCAGTGCCCA
30251 TGTCCCTTCC AGCTCACCTT GAGCAGCCTC AGATCATCTC TGTCACTCTG
30301 GAAGGGACAC CCCAGCCAGG GACGGAATGC CTGGTCTTGA GCAACCTCCC
```

FIGURE 3H

```
30351 ACTGCTGGAG TGCGAGTGGG AATCAGAGCC TCCTGAAGCC TCTGGGAACT
30401 CCTCCTGTGG CCACCACCAA AGGATGAGGA ATCTGAGTTG CCAACTTCAG
30451 GACGACACCT GGCTTGCCAC CCACAGTGCA CCACAGGCCA ACCTACGCCC
30501 TTCATCACTT GGTTCTGTTT TAATCGACTG GCCCCCTGTC CCACCTCTCC
30551 AGTGAGCCTC CTTCAACTCC TTGGTCCCCT GTTGTCTGGG TCAACATTTG
30601 CCGAGACGCC TTGGCTGGCA CCCTCTGGGG TCCCCCTTTT CTCCCAGGCA
30651 GGTCATCTTT TCTGGGAGAT GCTTCCCCTG CCATCCCCAA ATAGCTAGGA
30701 TCACACTCCA AGTATGGGCA GTGATGGCGC TCTGGGGGCC ACAGTGGGCT
30751 ATCTAGGTCC TCCCTCACCT GAGGCCCAGA GTGGACACAG CTGTTAATTT
30801 CCACTGGCTA TGCCACTTCA GAGTCTTTCA TGCCAGCGTT TGAGCTCCTC
30851 TGGGTAAAAT CTTCCCTTTG TTGACTGGCC TTCACAGCCA TGGCTGGTGA
30901 CAACAGAGGA TCGTTGAGAT TGAGCAGCGC TTGGTGATCT CTCAGCAAAC
30951 AACCCCTGCC CGTGGGCCAA TCTACTTGAA GTTACTCGGA CAAAGACCCC
31001 AAAGTGGGGC AACAACTCCA GAGAGGCTGT GGGAATCTTC AGAAGCCCCC
31051 CTGTAAGAGA CAGACATGAG AGACAAGCAT CTTCTTTCCC CCGCAAGTCC
31101 ATTTTATTTC CTTCTTGTGC TGCTCTGGAA GAGAGGCAGT AGCAAAGAGA
31151 TGAGCTCCTG GATGGCATTT TCCAGGGCAG GAGAAAGTAT GAGAGCCTCA
31201 GGAAACCCCA TCAAGGACCG AGTATGTGTC TGGTTCCTTG GGTGGGACGA
31251 TTCCTGACCA CACTGTCCAG CTCTTGCTCT CATTAAATGC TCTGTCTCCC
31301 GCGGAAAGCT CCACTGTGCT GCTGACTTGT CTCTGGTTTT CTGCAGTGTG
31351 GGGAGCCCAG GGAGGTGGAT GAATGAACAG TTAGTTACGC CCTGCCCACC
31401 TGCTGGGTGC CAGGCCTTCC TGTCCCTGTT GAATCCACTA GTTATCTCGT
31451 TCGATCTTTG CAGCAACCCT GTGAGATAGG AAGGTGTTAT TATCTTGCTT
31501 TGTCTTTCAA AAAAAAAGCG AGGCTCAGGG AGGCCAAGGG AAGTGTCCAA
31551 AGTCACACAT CAAGTTACTG GCAGTTACAG TTCCAACCAA GAGCTTCCAA
31601 CTCCATACCC CCTGCTCCTT CTGCTAGCCA TGAAGGGCTT TGGCCTTATA
31651 GGGCTTGTAG GGAAAGGTGA GTGGCCAAGA GCAAGTCCAT GCCAAGGGAA
31701 GATCTCCAAA CATGAGTCCC TGTCTGTTGC CTCCCCTGAG ATAGGCACAG
31751 GACAAGTGAT CAATGAGACA GGGTGGTCCT TGCCCTAAGA AGCAAAGTGT
31801 TTGGTTGGGG AGGGAAGTAG GGAAAAGGCT GCCACCTCCC CCCACCAAGG
31851 TACAACTGCT GACTTCCTTC CTCCCCAGCC CTCTATCACT GCCCTCTGTG
31901 CCGCTGCCGT TGACTGGCCT GCCCCACCAG ACTGAGGGCT CTGACTGCCC
31951 ACCGAGTCTA GTGTCAGCAT TATGGCTGAC CCAGAGCAGG CTATACAGTT
32001 AGTATGATGG ATAAATAAAT GATTGGTCAG TGCAGTCAAT TAGGTGCAAG
32051 CTGTTGGTAG TAGGCAAGGT CAATGAAGGT CATCCAAGGT GGGCATTGAA
32101 GGATGAGTAG AATGGCCAGG GGTAATGGGG GAGGAACTGG TGGGTGGGTG
32151 GAGGACTCTT CCAGACACCA TGTGGTTGAG GGCTGACAAA AAGCTGGGTG
32201 GAGGGCTTCC AGAGTGCCAA GCTCCCACCT GAAGAGGCTG ACCAGAGGCC
32251 AATCCTAAAC AACTCTAGGT GTTGGCTGGA GTTGCACTAA AGTGTATGGC
32301 CTCCCCAACC AAACCCTTTG CTTCTTAGGG CAAGGACCAC CCTGTCTCAT
32351 TGATCACTGT CCTGAGCCTA TCTCAGGGGA GGCAAAAGAG AGGGACCTGT
32401 ATTCAGAGAT CTTCCCTTGG CATGACTTGC TTTTGGCCAC TTACCTTTCC
32451 CTACAAGCTC TATGAGGCCA AGGCCCTTCA TGGTTAGTGT AAGGAGCAGT
32501 GGGCATGGAG TTGGAAGATC TGGGTTGGAA CGGTAACTGC CACTAACTCG
32551 ATGTGTGATT CTGAACACTT AACTTAGCCA TACATGCTCT CTTATTTGCT
32601 TTTGATGGCA AATAAGAGAA GGCCCAGCAA ACAGTGGCTT AAACCAGAAG
32651 GTCAATTAAT GTTTACTTTT CAGGAAGTCT GTAGGTAGAT GGTTGCTGGC
32701 ATTGGCCCAA CAGCTCATTT CAGCCTCCAA GGACTTGCGC TCCATAGTCC
32751 ACTCTGTCAT CTTAAAGCCT TCACACTTTT ACCCCCATGC TTGACCCCCA
32801 GGCTACATAC ACAGCT (SEQ ID NO:3)
```

FEATURES:

Start:	2697
Exon:	2697-2798
Intron:	2799-8874
Exon:	8875-9003
Intron:	9004-9252
Exon:	9253-9389
Intron:	9390-11975
Exon:	11976-12154
Intron:	12155-12893
Exon:	12894-13062
Intron:	13063-14905
Exon:	14906-15028
Intron:	15029-20092
Exon:	20093-20308
Intron:	20309-21836
Exon:	21837-21937
Intron:	21938-22861
Exon:	22862-22980
Intron:	22981-25083
Exon:	25084-25245
Intron:	25246-28357
Exon:	28358-28495
Intron:	28496-29686
Exon:	29687-29815

FIGURE 3I

Stop: 29816

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
609	T	G A	Beyond ORF(5')			
752	G	A	Beyond ORF(5')			
4623	A	- T	Intron			
4623	A	G T	Intron			
4699	C	T	Intron			
5062	A	G	Intron			
6158	T	C	Intron			
6573	C	A	Intron			
7120	A	G	Intron			
8411	A	C	Intron			
10035	A	G	Intron			
10849	G	A	Intron			
11916	T	C	Intron			
11962	C	T	Intron			
12333	C	T	Intron			
12375	A	C	Intron			
12418	T	C	Intron			
12603	G	A	Intron			
14225	G	C	Intron			
14416	C	T	Intron			
14643	A	C	Intron			
15612	C	T	Intron			
15685	C	T	Intron			
15971	C	T	Intron			
16175	G	A	Intron			
16589	C	A	Intron			
16980	A	-	Intron			
16993	T	- A	Intron			
17267	A	G T	Intron			
18804	C	T	Intron			
19084	C	T	Intron			
19669	G	A C T	Intron			
20397	G	C	Intron			
21575	T	C	Intron			
23363	C	T	Intron			
23413	T	C	Intron			
23945	T	G	Intron			
24483	C	A	Intron			
24643	A	G	Intron			
25329	G	A	Intron			
25421	G	A	Intron			
25797	T	G	Intron			
25926	C	T	Intron			
27289	T	C	Intron			
27591	G	T	Intron			
28245	T	A	Intron			
29337	G	A T	Intron			
29460	G	A C	Intron			
29994	A	T	Beyond ORF(3')			
30207	G	A	Beyond ORF(3')			
30497	T	G	Beyond ORF(3')			
30738	G	A	Beyond ORF(3')			
30758	T	C	Beyond ORF(3')			
31045	G	-	Beyond ORF(3')			
32591	C	T	Beyond ORF(3')			

Context:

DNA Position

609 ACACAGGAACAGAAAACCAAACACCACATGTTCTCAGTCATAAGTGGGAGTTGAACAGTG
AGAACGCATTGACACAGGGAGGGGAACATCACACACGGGGGCCTGTCAGGGGGTTGGAGG
GCAAGGGGAGGGAGAGCATTAGGACAAATACCTAATGCATGTGGGTCTTAAAACCTAAAT
GTCCGGTTGATAGCTGCAGCAAACCACCATGGCACATGTATACCTATGTAACAAACCTGC
ACATTCTGCACATGTATCCCAGAACTTAAAGTAAAATTAAAAAAAAAGAAAAGAAAAAAG
[T,G,A]
ACTGAAGTTGTTTACTTGCTCTCATTCATGCATCCCGGAGAAAAAGGTTTGAGTGCACAT

FIGURE 3J

```
        CCTGGATTAGGCACTGAGAAAGGCACTAGCTGGACAGGTGGTGATGAATAAAACAGACAG
        TAAATAGAAATTACATCATAATAATGTGTCATATATTTTAAAATAGCTACAAGATATTTT
        AAATGTTCTCACCACAAAGAAATGACAAATATTTGGGCCAGACGCGGTGGCTCACGCCTG
        TAATCCCAGCACTTTGGGAGACCGAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGAC

752     ACAAATACCTAATGCATGTGGGTCTTAAAACCTAAATGTCCGGTTGATAGCTGCAGCAAA
        CCACCATGGCACATGTATACCTATGTAACAAACCTGCACATTCTGCACATGTATCCCAGA
        ACTTAAAGTAAAATTAAAAAAAAAGAAAAGAAAAAAGAACTGAAGTTGTTTACTTGCTCT
        CATTCATGCATCCCGGAGAAAAAGGTTTGAGTGCACATCCTGGATTAGGCACTGAGAAAG
        GCACTAGCTGGACAGGTGGTGATGAATAAAACAGACAGTAAATAGAAATTACATCATAAT
        [G,A]
        ATGTGTCATATATTTTAAAATAGCTACAAGATATTTTAAATGTTCTCACCACAAAGAAAT
        GACAAATATTTGGGCCAGACGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGACC
        GAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCTAACATGGTGAAAC
        CCCATTTCTACTAAAAATGCAAAAAATTAGCCGGGCGTGGTGGTGCACACCTGTAATCCC
        AGCTACTTGGGAGGCTGAAGCAGGAGATTTGCTTGAACCTAGGTGGCAGAGGTTGCAGTG

4623    TCTCTCTCACACACACACACTCACAAACACACACAGACACACACACAAACACACAGACAC
        ACACAAAACACACGCACAAACACAAACACACACAAAAACACACGCACAAACACACACCCA
        AACACACAAACACAGACACACAAACACACACACAAACACACACAAAACACACAAAAAA
        CACAAACACACACATACACATACACACACACACACACACACACCCTGAACTGGAAACCCT
        AACTCAGTGTGTGTGTATGTGTGTGTATGTGTGTGTGTGTAAGAGAGAGAGAGAGAGAGA
        [A,-,T]
        TAAGCTGTCCTTTGAGTGAGGACCAGGGAGGGGAAGAAGAGAACCCAGGGAGAGTCCTTC
        CAAAGGCTGCCTTCACGAGCTTTCCTTCTGGCGGGGTTGGGTGAGGACCCTGGACCTTGT
        CTTCTTGTTTTTTCCCTTTCTGCCTGTTTTGGTCACCCTGCCCCCACCCTCCATGGCCGC
        CCCATTGTGCAAGGAAACCCAGAGGGTACACAGCACGGGCAGGGCAGCTGGGAAGCTGGT
        GAGAAGCTGGGAGGACCTTGGCAGCCTGAGCAACACAGTCCTTGCCAGGAGGTGACTCCC

4623    TCTCTCTCACACACACACACTCACAAACACACACAGACACACACACAAACACACAGACAC
        ACACAAAACACACGCACAAACACAAACACACACAAAAACACACGCACAAACACACACCCA
        AACACACAAACACAGACACACAAACACACACACAAACACACACAAAACACACAAAAAA
        CACAAACACACACATACACATACACACACACACACACACACACCCTGAACTGGAAACCCT
        AACTCAGTGTGTGTGTATGTGTGTGTATGTGTGTGTGTGTAAGAGAGAGAGAGAGAGAGA
        [A,G,T]
        TAAGCTGTCCTTTGAGTGAGGACCAGGGAGGGGAAGAAGAGAACCCAGGGAGAGTCCTTC
        CAAAGGCTGCCTTCACGAGCTTTCCTTCTGGCGGGGTTGGGTGAGGACCCTGGACCTTGT
        CTTCTTGTTTTTTCCCTTTCTGCCTGTTTTGGTCACCCTGCCCCCACCCTCCATGGCCGC
        CCCATTGTGCAAGGAAACCCAGAGGGTACACAGCACGGGCAGGGCAGCTGGGAAGCTGGT
        GAGAAGCTGGGAGGACCTTGGCAGCCTGAGCAACACAGTCCTTGCCAGGAGGTGACTCCC

4699    CAAACACAAACACACACAAAAACACACGCACAAACACACACCCAAACACACAAACACAGA
        CACACAAACACACACACAAACACACACACAAAACACACAAAAAACACAAACACACACATA
        CACATACACACACACACACACACACACCCTGAACTGGAAACCCTAACTCAGTGTGTGTGT
        ATGTGTGTGTATGTGTGTGTGTGTAAGAGAGAGAGAGAGAGAGAGATTAAGCTGTCCTTTGA
        GTGAGGACCAGGGAGGGGAAGAAGAGAACCCAGGGAGAGTCCTTCCAAAGGCTGCCTTCA
        [C,T]
        GAGCTTTCCTTCTGGCGGGGTTGGGTGAGGACCCTGGACCTTGTCTTCTTGTTTTTTCCC
        TTTCTGCCTGTTTTGGTCACCCTGCCCCCACCCTCCATGGCCGCCCCCATTGTGCAAGGAA
        ACCCAGAGGGTACACAGCACGGGCAGGGCAGCTGGGAAGCTGGTGAGAAGCTGGGAGGAC
        CTTGGCAGCCTGAGCAACACAGTCCTTGCCAGGAGGTGACTCCCAGGGCACGCCACCCTC
        TGCCAACACCCAGGCCTCTCTCCTCACCGACTGTCTCCAGTTTTCCTGTCTCCACCTGGA

5062    TCTGCCTGTTTTGGTCACCCTGCCCCCACCCTCCATGGCCGCCCCATTGTGCAAGGAAAC
        CCAGAGGGTACACAGCACGGGCAGGGCAGCTGGGAAGCTGGTGAGAAGCTGGGAGGACCT
        TGGCAGCCTGAGCAACACAGTCCTTGCCAGGAGGTGACTCCCAGGGCACGCCACCCTCTG
        CCAACACCCAGGCCTCTCTCCTCACCGACTGTCTCCAGTTTTCCTGTCTCCACCTGGATT
        CCCTCCTGGCCTCATCTCTGCTCCACTCTCTCTATCCTTCCTCTGGGTCTTTTTTTAATT
        [A,G]
        AAAAAAAATTTAATGAAATAAATGATAGATTTCTTGTATCACTTATTTTATTAAAATGTA
        AAAGGTTTCTTTTTTGCAAATCTGTAAGATATAAAGTAAAAATAAAAGTACACTCAAATC
        CCATAAGTTATTCACATTTTGATGAACATCTTTCCAGATGAATCTCTCTCTCTCTTCCCA
        GACACACACACACACACACACACACAGTAGGTTTTGCCTGCATTTTTTCATTAAGTGGTG
        TGTCAGGACACCCTGCCTTGTTAATGTTGAACTTTTCTAACATCCGCTTCCCATCCTGCC

6158    CCCGACACCAAGCCCCAAGGAAGTTAGTGGCTGCCAAAGGCCAGACAGTGGCTGACAGTG
        GGCGCCAATCATATCTGTCTGGTGTCAAAGCCTGGGCTTCCAGTCACGCTGCTGTTCCGC
        CTTTAGTTCAGCGGCTGTCAACCAGCATTGACCCTTTCTCCTGCCCTCACCCTGCCCCAC
        AACAGGGGAACTTTGGCAAAGTACAGAGACATTTTTTGCTGTCCAACCTGGAGACAGTCT
        TACTGGCATCTCATAGGTGGAGGCCAGCGGTGCTCTAAACACCCTGCAGTGCACAGCTCC
        [T,C]
        ACAACAAAGCATCATTTAGCCCAAAATGTCAGTGTGCCGAGGCTGAGGGACCCTGCCTCC
        CAGTAGGGAGGTGCCCTGGTTTGCTCGTGGGATGCTGAAAAAAGATTTATTTTTTTGTGG
        CTGATAACACAACCCTGACAAAGAATTTCCAAGTCTTCCTGCACTGTTTTGTGCAAATAA
        TACATACGCTCTTCTGGGTGATGAGAAGCAGGGATTGTGTACAGGTGCATCTGTTCTTCA
        GCAGCATTGTCAGAGTTAAACTCAGATGAATGCTATTGATTCCTTTAATAAACATTTGCA
```

FIGURE 3K

```
6573    TTGTGGCTGATAACACAACCCTGACAAAGAATTTCCAAGTCTTCCTGCACTGTTTTGTGC
        AAATAATACATACGCTCTTCTGGGTGATGAGAAGCAGGGATTGTGTACAGGTGCATCTGT
        TCTTCAGCAGCATTGTCAGAGTTAAACTCAGATGAATGCTATTGATTCCTTTAATAAACA
        TTTGCAAAGATGGCCGGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCG
        AGACAGGTGGATCACGAGGCCAGGAGATCAAGACCATCCTGGCCAACATGGTGAAATCCC
        [C,A]
        TCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCGCTTGTCTGTAGTCCCAGCTA
        CTCAGGAGACTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTAGAGGCTGCAGTGAGCCA
        AGATTGCACCACTGCACTCCAGCCTGGGGACAGAGCAAGGCTCTGTCTCAAAAATAAGTA
        AGTAAGTAAATAAATAAATAAATAAATAAATAAGCAAGAATTTGCAAAGATATCCTAAGT
        GTTGGGCCTGTTCTGGATGCTGAGGACGGTGATCTACAAATACAGCAGGTTCTTGAATAA

7120    CCTGTTCTGGATGCTGAGGACGGTGATCTACAAATACAGCAGGTTCTTGAATAATGTTGA
        TTCATTCAATATCATTTCATTATAATGTTGATGAGGGAGGGAAAAAAAAGGAAGGATCCC
        TTGAGCCCAGGAGATGGAGGTTACAGTGAGCTGTGACCGTGCCACTTCACTCCCACCTGG
        GCAACAGAGCCAGACCCTGTCTCAAAAAAAAAAAAAAAAGAAATAAAAGAGCGAGAGAGAA
        AGAAAAGAAAATGATTACTGGCTGGGGCCACTGTCTGTGTGGAGCGTGCACATTACCCTC
        [A,G]
        TGTCCACATGGCTTTTCTTTGGCTAGTATGGTTTCCTTCCACATCCCAAACCCGTGCACG
        TTAGGTGAATTGGAGTGTCTGTATGGTCCCTGTCTGAGTGAGCGTGGGCGTGCGTGTCAG
        TGTGCATTCTGCAATGGGATGGCATCTTGTCCAGGGCTGGTTTCCACCTTGTACCCTGAG
        CTGCCGGGACAGGATCTGGTCACCCAAGACCCTGACCTGCTGTAACTGGGTAAATAATTA
        TCTAACTTGTTTTCAATGTTTCTTAAGTATATGTATAGCTCACATTCCCTTCAGTGTTTA

8411    TTATTATTACTTTTATTTCAATAGCTTTAGGGGTACACGTAGTTTTCGGTTACATGAATG
        AATTGGATAGTGGTGAAGTCTGAGATTTTAATCCCTCCCTCCTATCCCACCCTGTCTGCT
        TCTAAGTCTCCAGTATCCATTCGACCACGCTATATACCTCTGGATACCCATAGCTTAGCT
        CCCACTTATAAGGGAGAACATGCACTATTTGGCTTTCCATTGCTGAGTCACTTCTCTTAG
        AACAATGGCCTCCTAGGCGGCAAGAGCGACACTCCATCTCAAAAATAAAATAATAATAAA
        [A,C]
        CCAAAAAAACCAGGTATTTTATTCTTCTTCTCCTTCTCCTCTTCCTCCTTTCTTTTCTTC
        CTTCTCCATCCCCCTTCCTCTCTTTCTTCTATCCCCTCCTCCTCCTTCTCCTCCTTCTCC
        TTCTTTCTCCTTTTCTTCCTTGTCCTATTCTTGATCTTTTCTTTTGAGAGGCAGCTAATC
        CAAGGTTTGAGAAGATGAAAGAACGTGCCTAGAACCACACAGCTGGGAAGGAGGGAGGCA
        GGGAGGAGGGGTGGGAATGGGGCAGGAGTCCTTTGCGAATAGATCCCTGGCCTGACCCGG

10035   AGGGAAGCAGGCCAGAAGTGGCTCAACTACCCAGCTCATGGGGAAGCAGAAAGGTCCTCT
        CTCCAAGCTGGAGCATCTATTCCCACTGCAAAGAAGCTTCTTATCTTCCCCGATATCACT
        CAGTACCCCAGCTTCTCTCTCCATTTCCAGGATCTCTCCTGCCAATCTAGCTAGCCATTT
        CCAGCTAAGCCATGGAGTCAATATAATCATAATCATAACCATAATCAATCATGATCATAA
        TGGGTATATTGAGTGTCTACAAGGCCCCAGGCATGATACCAGGAGCTTATGAATTGCCTC
        [A,G]
        TTTAATTCTTACCACAACTCTGAGAGCTGAGTATTCTTACTGCCCACATTCTGTGGATGA
        GGGATTGGAGGCAGAGAGGGATAAAGTGATTTGCTCATGGACACACGGGGATTGGACCCA
        GCTTCTCTGATAAGGCCTGTGTCCTCTCTAATCAGAAACTCAGGGCATATCTTCCTTTTG
        AGACAATGTGTCCCCTCAATGATGGCACGTCCTTGGCCCAGCCATCAGGAGTCAGCTGCT
        GGTCAGTTTACGGTAAATTCCTCCTGAGGCGCCCCTGTGTCAGGGGCTGTGCTAGACCCT

10849   GAATTCTTCTGCTTGGCTCCCTTGAGTGGCCAGCTTGGGTGGGAGGCCACTCCAGTGGGT
        TTCATTCTGCAGCATGCTGGAGAGCTTCCACTTCCAAACCCAAGTTCACACATGCTTCTG
        TATCCTTCCTGCCACCTTGCTCCTCTGAGTATGGTCTCCGGTTGTCCAAGGCACTGCCTG
        TCCTGGGAGTCACCTGTATGTGAGGCACCCTTGGTGCCTTGAGATATCATGTAGAAGCCT
        TGGTTCTTCTCAGACAACTCCATTCATGCAAACTCTCCCCCTCCTCCTAGCCTGGGTCCC
        [G,A]
        GGCTTTGTTTTTTTTTGGGTCCATAATGTCTGCCTGTGTGGACAGCAGCTTGGGCCCTGG
        TGCAGAACAGCTCCTAGGTCCCTTCTTCAGGCTCCTACCCCTGCCCCTGCTCCTACCCCC
        AGGTGAATTAGGAGCCCTGAGGAGGAGCCTGGCTGCAGCGAGGCCCACAGACTGAGAGTA
        GCTGAGCTCCTTCTGTCCCTAGCCTTGGACAGCTGGGGCATGTAGAGCCACAGAGCAGAG
        TCAGGCCCTGCCCTGCTCACAGCCCAAGGAGAGAGCAGACATGGAAACAGGTGCTTTGAA

11916   TCATATTGAGTTGGCACCATCTGGATGCCTGGGCTTCCCCTGCTAGATGGTGGGGCAGGG
        GTGCTCCTTAGAACCACGACTGGATCTGAGGCCTCTTGGTAACCCCAGAAGCAAGCAGAG
        TAGACATCAGTCATGGGTGTGGGAGAGGCAGGAGGGAGAGAGGAATGGAGGAAGCAAAGA
        AGGGAAGGAGGGAGGGAGGGGAGGCTCTAAAACCGTCATCCCTATTCCAATATCTGATCT
        TGAATTGGCCTCAACACCTGTGCATCCCTGCAGGGGTGGACCCAGTCCCCAGTTGCTTCC
        [T,C]
        AGGGAGTACGGGGGTGGGGTGGGGATTCTCTGGCTTTCCTCCCTGCCCCTCCTCTGCAGG
        CTGATGCTGGGCTTCATGGGCGTCACAGCCCTCCTGTCCATGTGGATCAGTAACACGGCA
        ACCACGGCCATGATGGTGCCCATCGTGGAGGCCATATTGCAGCAGATGGAAGCCACAAGC
        GCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGGCAAGGCCAAGGAGCTGCCAGGT
        GAGCCCCTGGCCAGGGCACTGCCAGGCCACAACAGCAGCCTTCCCCTCCCTCTGCTGGCA

11962   ATGGTGGGGCAGGGGTGCTCCTTAGAACCACGACTGGATCTGAGGCCTCTTGGTAACCCC
        AGAAGCAAGCAGAGTAGACATCAGTCATGGGTGTGGGAGAGGCAGGAGGGAGAGAGGAAT
        GGAGGAAGCAAAGAAGGGAAGGAGGGAGGGAGGGGAGGCTCTAAAACCGTCATCCCTATT
```

FIGURE 3L

```
        CCAATATCTGATCTTGAATTGGCCTCAACACCTGTGCATCCCTGCAGGGGTGGACCCAGT
        CCCCAGTTGCTTCCCAGGGAGTACGGGGGTGGGGTGGGGATTCTCTGGCTTTCCTCCCTG
        [C,T]
        CCCTCCTCTGCAGGCTGATGCTGGGCTTCATGGGCGTCACAGCCCTCCTGTCCATGTGGA
        TCAGTAACACGGCAACCACGGCCATGATGGTGCCCATCGTGGAGGCCATATTGCAGCAGA
        TGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGGCAAGGCCA
        AGGAGCTGCCAGGTGAGCCCCTGGCCAGGGCACTGCCAGGCCACAACAGCAGCCTTCCCC
        TCCCTCTGCTGGCAAATGCTTTGGCCACCTCCTTCTCCCTGTCTGCTTCCCGGAGCCCTC

12333   GGCAACCACGGCCATGATGGTGCCCATCGTGGAGGCCATATTGCAGCAGATGGAAGCCAC
        AAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGGCAAGGCCAAGGAGCTGCC
        AGGTGAGCCCCTGGCCAGGGCACTGCCAGGCCACAACAGCAGCCTTCCCCTCCCTCTGCT
        GGCAAATGCTTTGGCCACCTCCTTCTCCCTGTCTGCTTCCCGGAGCCCTCCTTTAAACAC
        GCATAGAGAAAAAAAAATAGAAAATACTGTTGTCCTAAGTTTTAGGAGGGGATTATTGCA
        [C,T]
        ACAACTTAGATCCTTTAATAGAGCTTTGAACAAAGTCTCACCCTCAGTTCCCATCAGTTG
        CAGAAATCAGTGTGTTCACCTGATTATTCATTTGGGCATCTTTCGAGCACTTAGGGATGC
        CCCTCACTCCTTGCTACTCCTGCTCATCCTCAAGGAGGCCTTTTCTGACCTCCTCGAGCA
        GCTCAAATCCTTCCACTCTCTGCTCCCATAGGTCTGGGGCTTGGCGTCCCATGCTTGCTT
        CCCTGCTAGGTGCGAAGCTCAGGGAAGACGAGTCAGCATCTACCTTGCCGTCTGCCGTGT

12375   GCAGCAGATGGAAGCCACAAGCGCAGCCACCGAGGCCGGCCTGGAGCTGGTGGACAAGGG
        CAAGGCCAAGGAGCTGCCAGGTGAGCCCCTGGCCAGGGCACTGCCAGGCCACAACAGCAG
        CCTTCCCCTCCCTCTGCTGGCAAATGCTTTGGCCACCTCCTTCTCCCTGTCTGCTTCCCG
        GAGCCCTCCTTTAAACACGCATAGAGAAAAAAAAAATAGAAAATACTGTTGTCCTAAGTTT
        TAGGAGGGGATTATTGCACACAACTTAGATCCTTTAATAGAGCTTTGAACAAAGTCTCAC
        [A,C]
        CTCAGTTCCCATCAGTTGCAGAAATCAGTGTGTTCACCTGATTATTCATTTGGGCATCTT
        TCGAGCACTTAGGGATGCCCCTCACTCCTTGCTACTCCTGCTCATCCTCAAGGAGGCCTT
        TTCTGACCTCCTCGAGCAGCTCAAATCCTTCCACTCTCTGCTCCCATAGGTCTGGGGCTT
        GGCGTCCCATGCTTGCTTCCCTGCTAGGTGCGAAGCTCAGGGAAGACGAGTCAGCATCTA
        CCTTGCCGTCTGCCGTGTTCCCTTACCATCCCCAGCCCAGTGCAGTAGAGTCAGGGTCTG

12418   GAGCTGGTGGACAAGGGCAAGGCCAAGGAGCTGCCAGGTGAGCCCCTGGCCAGGGCACTG
        CCAGGCCACAACAGCAGCCTTCCCCTCCCTCTGCTGGCAAATGCTTTGGCCACCTCCTTC
        TCCCTGTCTGCTTCCCGGAGCCCTCCTTTAAACACGCATAGAGAAAAAAAAATAGAAAAT
        ACTGTTGTCCTAAGTTTTAGGAGGGGATTATTGCACACAACTTAGATCCTTTAATAGAGC
        TTTGAACAAAGTCTCACCCTCAGTTCCCATCAGTTGCAGAAATCAGTGTGTTCACCTGAT
        [T,C]
        ATTCATTTGGGCATCTTTCGAGCACTTAGGGATGCCCCTCACTCCTTGCTACTCCTGCTC
        ATCCTCAAGGAGGCCTTTTCTGACCTCCTCGAGCAGCTCAAATCCTTCCACTCTCTGCTC
        CCATAGGTCTGGGGCTTGGCGTCCCATGCTTGCTTCCCTGCTAGGTGCGAAGCTCAGGGA
        AGACGAGTCAGCATCTACCTTGCCGTCTGCCGTGTTCCCTTACCATCCCCAGCCCAGTGC
        AGTAGAGTCAGGGTCTGTGGCTGACGGCCTGATTGCCAGACCCTGGGCAAGGTCCTGGGG

12603   TGTCCTAAGTTTTAGGAGGGGATTATTGCACACAACTTAGATCCTTTAATAGAGCTTTGA
        ACAAAGTCTCACCCTCAGTTCCCATCAGTTGCAGAAATCAGTGTGTTCACCTGATTATTC
        ATTTGGGCATCTTTCGAGCACTTAGGGATGCCCCTCACTCCTTGCTACTCCTGCTCATCC
        TCAAGGAGGCCTTTTCTGACCTCCTCGAGCAGCTCAAATCCTTCCACTCTCTGCTCCCAT
        AGGTCTGGGGCTTGGCGTCCCATGCTTGCTTCCCTGCTAGGTGCGAAGCTCAGGGAAGAC
        [G,A]
        AGTCAGCATCTACCTTGCCGTCTGCCGTGTTCCCTTACCATCCCCAGCCCAGTGCAGTAG
        AGTCAGGGTCTGTGGCTGACGGCCTGATTGCCAGACCCTGGGCAAGGTCCTGGGGCTTAC
        AGAGAGGAATCGGGCACATCCCTGCCAGCAACTCTTATGGAGCCCAGTGGGGCAGCTAAA
        TCAGCAGAGCTGGGATTTCCCAATCCTCAGGTCAGCAGCAGAGTCAGGACCTGGGGCTGG
        GTGGGCAGCCCCCATGACTGGCTCAGCTAACAGCGCTGTGCCCACCACAGGGAGTCAAGT

14225   GGCTGGGCAGAAACGTGAGGTTCAACAACCCGTTTGTTTTAATTTCGGGAGTGTTTTCTG
        TAATGATATCCTTACAGTTCTCCAGTAACTTTCTTTGGGAAGAGCAGCCCGTCTGGGCTG
        AGTGGGGAAAGCTCTGCGCCTGCTTTGACACTCTTGAGCTAAAGGGGGCGCCCCTGGGGC
        TAGCAGAGCCCCGGGGATGGGAGGCGGGGCCTGTGGTGGAAGTGACCCTCCTCCAGCCTC
        CGCTCTGGGAAGCTTTTGAGATTTCCTTTGCTAAGTGGGGGGACCGTTCTTTGCAGAAAC
        [G,C]
        CACAGAGCGAGATTGCTGAGGTCTCTGCAGATCCCCAAAGATGTCAGCCAAATTACATGC
        ATGTGTATAAAAGGTGTATTTTTCTTTTTTTTCTTTTTGAGACAAGTCTCGCTCTGTCGC
        CCAGGCTGGAGTGCAGTGGCGCGATGTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAA
        GCGATTCTCCCGCTTCAGCCTCCCTATTAGCTGGGATTACAGGCGCCCGCCACCATGCCT
        GGTTAATTTTTGTATTTTTAGTGGAGACGGGGTTTCACCATGTTGGCCAGGCCAGTCTTA

14416   CGGGGATGGGAGGCGGGGCCTGTGGTGGAAGTGACCCTCCTCCAGCCTCCGCTCTGGGAA
        GCTTTTGAGATTTCCTTTGCTAAGTGGGGGGACCGTTCTTTGCAGAAACCCACAGAGCGA
        GATTGCTGAGGTCTCTGCAGATCCCCAAAGATGTCAGCCAAATTACATGCATGTGTATAA
        AAGGTGTATTTTTCTTTTTTTTCTTTTTGAGACAAGTCTCGCTCTGTCGCCCAGGCTGGA
        GTGCAGTGGCGCGATGTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTCC
        [C,T]
        GCTTCAGCCTCCCTATTAGCTGGGATTACAGGCGCCCGCCACCATGCCTGGTTAATTTTT
```

FIGURE 3M

GTATTTTTAGTGGAGACGGGGTTTCACCATGTTGGCCAGGCCAGTCTTAAGCTCCTGACC
TTGTGCCCCACCTGCCTCGGCCTCCCAGAGTCCTGGAATTACAGGCGTGAGCCCCTGCGC
CCGGCCACAAAGTTGTATTTTTCTGGAGGGATGGGCCATAACTTCCATGAGACTCTTAGC
AAGGCCTGGACACACAGAAGAGTCAGTGGGTCATTTCTCGGCCTTGTCTTGTGCTGTGGC

14643 CGCCCAGGCTGGAGTGCAGTGGCGCGATGTTGGCTCACTGCAACCTCTGCCTCCTGGGTT
CAAGCGATTCTCCCGCTTCAGCCTCCCTATTAGCTGGGATTACAGGCGCCCGCCACCATG
CCTGGTTAATTTTTGTATTTTTAGTGGAGACGGGGTTTCACCATGTTGGCCAGGCCAGTC
TTAAGCTCCTGACCTTGTGCCCCACCTGCCTCGGCCTCCCAGAGTCCTGGAATTACAGGC
GTGAGCCCCTGCGCCCGGCCACAAAGTTGTATTTTTCTGGAGGGATGGGCCATAACTTCC
[A,C]
TGAGACTCTTAGCAAGGCCTGGACACACAGAAGAGTCAGTGGGTCATTTCTCGGCCTTGT
CTTGTGCTGTGGCCATGTTCTGAGGCTCCCACTCGATTAGGGGACAATGCTTGGCAATGG
ACTTGGTGGCTAGACCTCAGGAGGATGTGGCCTCCACACAGGCGCGCCTCTCAGGGCCCA
GCTGCTGCTCCGTCCCCACGCACAGGGCCAGGCTGGCTCCCACAGCTCAGCATCTGAGGT
GGGGGCCGGTGTCTTCTTGTAGGTTGTTTCCTGACAGCAAGGACCTCGTGAACTTTGCTT

15612 CCCATTCCATCTCTGAGCGGCCCCTGGGCATATCACAGGCCTGTCCTTTAGTATCTGCAT
TTGGCTTCCGGTGACTTTGAATTCCTCCAGAACCACTCTGATGCTGGGCACCCCGCACAG
CTCCCAGCACAGGGAGGAAGAGCAGGCAGGTTAAAGCAATTAAAGATAAGCTGGTCCCCA
CGTGCCAGTTCGACATTGCTGGACAAGCTTCCTCTTTGCCGTGTGGGTCCATCAGGCCAG
GTCACCGCAAACCTGTGACTTAGCTCTGAGCTGAGCGCATACGCTCTGTGCCTCAATGCA
[C,T]
GGGGAGTTTAAGTCGAGTAAAACCAGCAGTGATTATGACCAAATCCATCCAAACCCAGAC
ATTTACTGAATACCTCTGGTGTTCCCAGCAGTGTACAGGTCCTAGAAAGTTTACCTTCCT
GTTCCTAGCACACAGGCAAGTTCATCAGGGGTCACCTTTGATGGCAGCCAGACTTTGGAC
AGAAACCATGACCTGTGGCTGACAAATAGCTAAAAAAAAGTTATTGTTTTTCTAAAACAC
ACAAATTTATCTGTGGTGCAAAGGTGATCAGGCCACACCAGGATAGAAAGTACTCAGCTC

15685 ACTTTGAATTCCTCCAGAACCACTCTGATGCTGGGCACCCCGCACAGCTCCCAGCACAGG
GAGGAAGAGCAGGCAGGTTAAAGCAATTAAAGATAAGCTGGTCCCCACGTGCCAGTTCGA
CATTGCTGGACAAGCTTCCTCTTTGCCGTGTGGGTCCATCAGGCCAGGTCACCGCAAACC
TGTGACTTAGCTCTGAGCTGAGCGCATACGCTCTGTGCCTCAATGCACGGGGAGTTTAAG
TCGAGTAAAACCAGCAGTGATTATGACCAAATCCATCCAAACCCAGACATTTACTGAATA
[C,T]
CTCTGGTGTTCCCAGCAGTGTACAGGTCCTAGAAAGTTTACCTTCCTGTTCCTAGCACAC
AGGCAAGTTCATCAGGGGTCACCTTTGATGGCAGCCAGACTTTGGACAGAAACCATGACC
TGTGGCTGACAAATAGCTAAAAAAAAGTTATTGTTTTTCTAAAACACACAAATTTATCTG
TGGTGCAAAGGTGATCAGGCCACACCAGGATAGAAAGTACTCAGCTCTGAGTTAAGTGCC
TGTGCTCTGTGCCTCCATCCACAGGAAGTTCGAGCCAAGTCAAACCAGGGGAATTTGTGA

15971 ACATTTACTGAATACCTCTGGTGTTCCCAGCAGTGTACAGGTCCTAGAAAGTTTACCTTC
CTGTTCCTAGCACACAGGCAAGTTCATCAGGGGTCACCTTTGATGGCAGCCAGACTTTGG
ACAGAAACCATGACCTGTGGCTGACAAATAGCTAAAAAAAAGTTATTGTTTTTCTAAAAC
ACACAAATTTATCTGTGGTGCAAAGGTGATCAGGCCACACCAGGATAGAAAGTACTCAGC
TCTGAGTTAAGTGCCTGTGCTCTGTGCCTCCATCCACAGGAAGTTCGAGCCAAGTCAAAC
[C,T]
AGGGGAATTTGTGACCAGAGGGAAGAGACTGCAGAGCTCAGAGGCAAAAGTGCCCACGGA
AACCTGTGATTTTGTGGGGAAAATAGGGAATTTTCCTAAGTTTTCTTCTGAAGGAGGAAC
TGTTTTGAAAACTCCCATTAAAAAGTTGCTATACAGGCCGGGCGCGATGGCTCACACCTG
TAATCCCAACATTTTTGGAGGCCGAGGTGGGCAGATCGCCTGAGGTCAGGAGTTTGTAAC
CAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCGGGCGT

16175 GGTGATCAGGCCACACCAGGATAGAAAGTACTCAGCTCTGAGTTAAGTGCCTGTGCTCTG
TGCCTCCATCCACAGGAAGTTCGAGCCAAGTCAAACCAGGGGAATTTGTGACCAGAGGGA
AGAGACTGCAGAGCTCAGAGGCAAAAGTGCCCACGGAAACCTGTGATTTTGTGGGGAAAA
TAGGGAATTTTCCTAAGTTTTCTTCTGAAGGAGGAACTGTTTTGAAAACTCCCATTAAAA
AGTTGCTATACAGGCCGGGCGCGATGGCTCACACCTGTAATCCCAACATTTTTGGAGGCC
[G,A]
AGGTGGGCAGATCGCCTGAGGTCAGGAGTTTGTAACCAGCCTGGCCAACATGGTGAAACC
CCGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTAGCCCACGCCTGAAATCCCAG
CACTTTGGGAGGCCAAGGAGGGCGGATCGCCTGAGGTCAGGAGCTCGAGACCAGCCTGGC
CAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAGTTAGCTGGGCATGGTGGCACA
TGCCTGTAACCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAGCCGGGAGGT

16589 ATCCCAGCACTTTGGGAGGCCAAGGAGGGCGGATCGCCTGAGGTCAGGAGCTCGAGACCA
GCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAGTTAGCTGGGCATGG
TGGCACATGCCTGTAACCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAAGCC
GGGAGGTAGAGGTTGCAGTAAGCCAAGATCATGCCACTGCACTCCAGCCTGGGCGACAGA
GCAAGACTCTGTCTCAAAACAAAAAAAAAAAGTTGCTATACATATTCAAAACAATCATAAT
[C,A]
ATGATAGTAAGAATGACAATATTAATGATCATTGCCCAAACCCCACTCTGTCCTGCCCAT
GGACGGGGCAGGGGAAACTGTTTGCATGGCTGCCTGGCCACCCAGCCTGGCTTTGACAGT
AGCTCTCTTTGCCCTGCCTCTTGAATCTGCACCAGGGCCAAAGTCCTGTTCATTTGTTCA
CATCCGTCGAACAGGTCTCTCAGGAGATGGTCCTGAACCTGCTGCAGGTGAGCATCTGTG
TCTCCTCATGGGGCAACAGGAATAATAATGACCAACATTTATTGAGTGCTCATCATGTGC

FIGURE 3N

16980 TGCCTGGCCACCCAGCCTGGCTTTGACAGTAGCTCTCTTTGCCCTGCCTCTTGAATCTGC
ACCAGGGCCAAAGTCCTGTTCATTTGTTCACATCCGTCGAACAGGTCTCTCAGGAGATGG
TCCTGAACCTGCTGCAGGTGAGCATCTGTGTCTCCTCATGGGGCAACAGGAATAATAATG
ACCAACATTTATTGAGTGCTCATCATGTGCCAGACATGATTTCGAGCGCTCTTTTCCTTT
CTTTATTTTATTTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT
[A,-]
TTTATTTTTGAGACAGTGTCTTGCTCTGTCACCCATGCTGGAGTGCAGTGGTATGATCTC
GGCTCGCTGCAACCTCCACCACCTGGGTTCAAGCAATTCCCCCTGCCTCAGCCTCCCAAG
TAGCTGGAATTACAGGCACCCACCACCACCATGCCTGGCTAATTTTTGTATTTTTTAGTA
GAGATGGGGTTTTGCCATGTTGGCCAGGCTGGTCTTGAACTCCTAACCTCCGGTGATCCG
CCCTCCTTGGCCTCCCAAAGTGCTGGCGTTACAGATGTGAGCCACCTCGCCTGGCCCAAG

16993 AGCCTGGCTTTGACAGTAGCTCTCTTTGCCCTGCCTCTTGAATCTGCACCAGGGCCAAAG
TCCTGTTCATTTGTTCACATCCGTCGAACAGGTCTCTCAGGAGATGGTCCTGAACCTGCT
GCAGGTGAGCATCTGTGTCTCCTCATGGGGCAACAGGAATAATAATGACCAACATTTATT
GAGTGCTCATCATGTGCCAGACATGATTTCGAGCGCTCTTTTCCTTTCTTTATTTTATTT
TATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTTTTGAG
[T,-,A]
CAGTGTCTTGCTCTGTCACCCATGCTGGAGTGCAGTGGTATGATCTCGGCTCGCTGCAAC
CTCCACCACCTGGGTTCAAGCAATTCCCCCTGCCTCAGCCTCCCAAGTAGCTGGAATTAC
AGGCACCCACCACCACCATGCCTGGCTAATTTTTGTATTTTTTAGTAGAGATGGGGTTTT
GCCATGTTGGCCAGGCTGGTCTTGAACTCCTAACCTCCGGTGATCCGCCCTCCTTGGCCT
CCCAAAGTGCTGGCGTTACAGATGTGAGCCACCTCGCCTGGCCCAAGCACTCTTAAACTT

17267 TATTTATTTATTTATTTATTTTTGAGACAGTGTCTTGCTCTGTCACCCATGCTGGAGTGC
AGTGGTATGATCTCGGCTCGCTGCAACCTCCACCACCTGGGTTCAAGCAATTCCCCCTGC
CTCAGCCTCCCAAGTAGCTGGAATTACAGGCACCCACCACCACCATGCCTGGCTAATTTT
TGTATTTTTTAGTAGAGATGGGGTTTTGCCATGTTGGCCAGGCTGGTCTTGAACTCCTAA
CCTCCGGTGATCCGCCCTCCTTGGCCTCCCAAAGTGCTGGCGTTACAGATGTGAGCCACC
[A,G,T]
CGCCTGGCCCAAGCACTCTTAAACTTAATTAATTTTCACAACAACCTGCGAGGTCAGCAC
TATTATTATTATTCCCAATTTACAGACAAGGCAACTGAGGCATGGAGAGGTGATGTGGTC
AACACAGAGCTTTGTAACAGGGAAGTAGGGGGACTGAGACTTGAACCCAGGCCCTTTGGC
TCCCACTGCATGGCATCCCCTCTTGGGGAGGCTGAGGGTTGCTGTCCTTAGTTGCCTCCA
GACCTAAGCATGACCAGGTGTCAGAAACACTAGTTGGGGCCGGGGCTGCCCTAGAACCCC

18804 AACTGGATCCATTTACTTGCATCTCATGCCAGCCTGGTCATCACCAGATGAAATTAACCC
AGAGATGAGAGCAAAGCTGCTCAGCACGAGAGACTCTGAAGGCTTGGCGGTACCACTGTG
GGGCACTGGCATTGGAAGACTGCATACTCCATGCAGCCCCAGAGTCTGCAGCTACTGTGG
TGTTGGGGATGAGCTGCCAGCACCAAATGCAGGCTCTGGCTCCTGGGCCACTAGTAATAC
CAAGGTCACCCCTTATGCTGGAAACCTGAAGCCCCTGGCTGAGCCCCAGGGTCTCTAGGA
[C,T]
GACAGTTGGCAGCAGAGAGGTGCTTGGTAGAGCACAAACTTTACTAAGCCAAGGGTGTGG
CAGCAGAGAGGCCCTGTCTTACACCAGCAGAGCCATCCCTGTGCCGGATGTCTAGAGAGT
GTCCCTAGCGGGTGACCCTCAGGACACACGGCCTTGCCCAGCAGGGAGATCCTAGCCAGC
CGTGTAGACCTGAGGTCCCATCAGTGTTGCCTCCTTTTCTGACCCCTGAGCACCCCAGAA
AGCTGTGACCTGATGTCCTGGTGTCCCCATGTTCCAGGCCAAGCCACCATCACACCAACA

19084 GAGCCCCAGGGTCTCTAGGACGACAGTTGGCAGCAGAGAGGTGCTTGGTAGAGCACAAAC
TTTACTAAGCCAAGGGTGTGGCAGCAGAGAGGCCCTGTCTTACACCAGCAGAGCCATCCC
TGTGCCGGATGTCTAGAGAGTGTCCCTAGCGGGTGACCCTCAGGACACACGGCCTTGCCC
AGCAGGGAGATCCTAGCCAGCCGTGTAGACCTGAGGTCCCATCAGTGTTGCCTCCTTTTC
TGACCCCTGAGCACCCCAGAAAGCTGTGACCTGATGTCCTGGTGTCCCCATGTTCCAGGC
[C,T]
AAGCCACCATCACACCAACACTTGGCCCTCACACTCTCCAAGGCTGTTCACATCCAGCAC
TGGCTTCAGGAATGAGCTCCTATTCCATCAACCCCTTCCCTCCTATGATTATGTCTCATG
GCCCCCGGGAAGGGCTCTCACGAGGGAGGGCTCTCCAGGACAATACTCTTGGCCTTGCCC
ACCCCTTCAAACCAACAGTGGCTGGAACTGGAATGTGTGAATGGAATATTCAGCATACCT
TGAGGCCTTAGTCCTATGCACAGTGGCCCCAGTTATCCCCCCTCCACAGCTGAGCTCCCC

19669 CACAGCTGAGCTCCCCTTTACACCTCCTCCAAGAACCTCCTCTCCTCCCTGCCTCCTCAT
GCCAACGCCACCTTAGGGGAGGCCCTGCAGGACACCCTGGACAATGGACACTGGTCCCAA
GGGGGCCCATCCAGGATGGGGGTGCCATCCTGGGCTGTCTTCCTTCTTGCCCTAGCCATG
CTTGCTGCTAACCCCAGGGTCTCCTGGATCCCTAATCCTGCACCTCCAACTCCAGGGAAC
ACAAGGACCCATTCTGCCCCTGACTAGCCCTGTCTGCCAGGGTTCATACTCACTCCCTGC
[G,A,C,T]
TCTCCCTGAGCCACCTTGGTGATGGGGGTTGGCATCCCAACACCATCGAAGGCAGCTCCA
GGCTGAGGTGGAAGGAGGAAGACTTGGGAAGCATGTGAGGGAGCCCTGTTCCCACCTTGC
GCAGGCTCCGAAGCTCCTTATGGCCTTCCCCCAGGTGACCCTGGAGCAGCCAGTCTCCAG
GTGTCTGGGCACCTGCCGAGACCCTCTAGCCTCTCTACAGAGACTTTTTCCCTAGTACAT
TCTGGGATGGAAGAACAGGAGAGGGAAAGAGGCAGGAAGGGCCTTTCTCCAGGCCCCATA

20397 AAAAAGTCCTGGGGCTGCGGGCTAGAGAGCAAGAAAAACGAGAAGGCTGCCCTCAAGGTG
CTGCAGGAGGAGTACCGGAAGCTGGGGCCCTTGTCCTTCGCGGAGATCAACGTGCTGATC
TGCTTCTTCCTGCTGGTCATCCTGTGGTTCTCCCGAGACCCCGGCTTCATGCCCGGCTGG

FIGURE 3O

```
        CTGACTGTTGCCTGGGTGGAGGGTGAGACAAAGTAAGTCTTGGATTCAATAGAAATCGCT
        GGCTTAGGGCCAGGCGCGTTGGCTCACACATGTAATCCCAGCACTTTGGGAGGCTGAGGT
        [G,C]
        GGTGGGTCACTTGAGGTCAGGAGATCGAGACCATCCTGGCCAACATGGTGAAACCCTGTC
        TCTACTAAAAATACAGAAAATTAGCGAGGCATGGTGGCACATGCCTGTAGTCCCAGCTAC
        TTGGGAGACTGAGGCAGGAGAATCACTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCCA
        GATCGTGCCACTGCACTCCAGCCTGGGCAACAGAGAGAGACTCCGTCTCAAAAAAAGAGA
        AAGAAAGACACCACTGGCTTAGTGCACTAGTGCCTAAATGCTGCTGGTCTCGGCTACAGG

21575   AGGTCCTGAAATACACTTATGGAGAATGCACGCTGAGAGGGGGAAGTAAACTGCTTAGGA
        TCACCCAAAGTTGGTGGTCAAGAGTGTGGGCATCTTGATTTCTAGCCAGGATTCAGTCTC
        CCATACCACTCTTATTTTTTTATTTTTTTGAGACAGAGTCTCACACTGTCACCCAGGCTG
        GAGTGCAATGGCATGATCTCAACTCACTGCAACCTCCACCTCCCAGCAATTCTCCTGCCT
        CAGCCTGCCGAGTAGCTGGGATTACAGGCGCCCGCCAGCATGTCTGGCTAATTTTTTGTA
        [T,C]
        TTTTAGTAGAGACGGGGTTTCACTATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCGT
        GATCCGCCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAAGTGTGAGCCACTGCACCTGG
        CCACCACTCTTGACCTTGACTTTTAAGGCTGTGAGCCTGTTTCTTTGCATAGAAGCATTT
        GGACACAGAACTGCCGGAGTTGTGATGGGTTTGTTGAGTGACTGTCTCTGTCGCAGATGA
        GCTGTGCTTTTCCCCACCTAGGTATGTCTCCGATGCCACTGTGGCCATCTTTGTGGCCAC

23363   TGTGTGTTTCTGTGCCTGCATCCTTCGTATAACCGCACATTCCTTGAGGACATGGACTCT
        GTCTTGTCATCTAGGAACTCTACACCACACACAGGGCCTGGAAGACAGAAAGTAACCTTT
        GAGCGATTGCAGGAATGAGTGAATGAGTGACCGTGGTTAGCCAAGAGAGGCAGAGGACAC
        TGTCAGTTACCCTCTGGGGCTTGATCACAATAATCTCTGCTTTGATTTGTCTGAGGGAAA
        TCTTTCTTTCCAATCCTTGTCAATATTGTTTGCTACTACTTTTGGTCCTTCTACTGGCTA
        [C,T]
        TTAACATGGTAGCTACTTCAAAATTTTTCTTTAGCTAAGTATGTAGCAGCGTAGGAGGTG
        AGGAACATGTTGGAAAACACACAAAAATATAACTTTCTTTACCTCCTTCTTTCCCTCCTG
        GGGAAGAAATGAGCCAGAGGGAGGGATGAGCTAGCTTGCTGCTGCTGTCCTCCAACCAAC
        CATCTACCTACCCAAGTATCCAGGAGTGTAATAGACAGACTTGGTCTAGTTATTGCTGTT
        TCTTCAATATCTAGGACACAGCCTGGTGCCTAGTGGGTGCTAAGTTTTTGCGGAGGTGAA

23413   CATGGACTCTGTCTTGTCATCTAGGAACTCTACACCACACACAGGGCCTGGAAGACAGAA
        AGTAACCTTTGAGCGATTGCAGGAATGAGTGAATGAGTGACCGTGGTTAGCCAAGAGAGG
        CAGAGGACACTGTCAGTTACCCTCTGGGGCTTGATCACAATAATCTCTGCTTTGATTTGT
        CTGAGGGAAATCTTTCTTTCCAATCCTTGTCAATATTGTTTGCTACTACTTTTGGTCCTT
        CTACTGGCTACTTAACATGGTAGCTACTTCAAAATTTTTCTTTAGCTAAGTATGTAGCAG
        [T,C]
        GTAGGAGGTGAGGAACATGTTGGAAAACACACAAAAATATAACTTTCTTTACCTCCTTCT
        TTCCCTCCTGGGGAAGAAATGAGCCAGAGGGAGGGATGAGCTAGCTTGCTGCTGCTGTCC
        TCCAACCAACCATCTACCTACCCAAGTATCCAGGAGTGTAATAGACAGACTTGGTCTAGT
        TATTGCTGTTTCTTCAATATCTAGGACACAGCCTGGTGCCTAGTGGGTGCTAAGTTTTTG
        CGGAGGTGAACAAATCCATCCATCTAGTCACCTCTCCATCCATCATCCATGCATCTATTC

23945   AAGTTTTTGCGGAGGTGAACAAATCCATCCATCTAGTCACCTCTCCATCCATCATCCATG
        CATCTATTCATGTCTTCATCCATCCATCCGTCCGTCCATCCGTCCATCCATCCATCCATC
        CATCCATCCATCCATCCATCCATCTGTCCATCCATCCACCCATCTATCTATCCACC
        CATCTATCTATCCATCCATCCATCCACCCATCTATCCATCCATCCATCCATCCATC
        CATCCATCCATCCATCTATCCATCCATCCATCCAACATTCCTATTATGTCCCTAGTGTTA
        [T,G]
        GCCAGGCACAGAGATTACAGAGGAGATTGAGATACGGTCCCTGTTCGTGGCAGACTTCAC
        AGACTAGGGAGGGGCACATATATGAAAGGGCATTTCAGGAAGTAGCACACGAGCAAGGGA
        AAAATGTGAGGTATTTAGCTGAGGAGAAGTAGAAGATGAGGCTGGTAAGGCTACCAGAAG
        CCACTTCTCTGAGGGCCCCAAGATAGAGGGGTGTGGACTTGATCGTGAATGCAGTAGACA
        GCCACTGAAGGACTGAGGCCAGGGGGTGAGTTGGTCAGATCTGCACATGAGGAAATCACT

24483   ACAGCCACTGAAGGACTGAGGCCAGGGGGTGAGTTGGTCAGATCTGCACATGAGGAAATC
        ACTCTGATGTCTGGAGTGGGGGCCTGGGCTGGGCAGGGCTTGGAGGAGAACTAGCTGAGA
        CTCTGCAGCCTTCCATCTCACTCAGGCTCAGAACTTTGGACTCTGTGGACATTCTCTCCT
        CCTTTGGCCCCCAGCTCAGCACAGTCTCCAGCTTTACTTCGGACTCAGACTATTCCTGCT
        CAGCCTTCGTTGCTGACTTCTCTGTTCTCCCTGAAACAGGAGTGTCTGCCCAGGCTCTGT
        [C,A]
        CTTGGCCTCTCCTCTTTTTACACTTCATTCTCTCCCTGGACAATCTCTTCTCAGCCCAAA
        GCCCTAAATCTAAACCTTCAATTTCTGGTTGAAATCATTCTCCTGAGCTTCCAAAACTGT
        GGAGCACTGAAGAGGAGGAGATGGATGTGAGACATTTGGGTGACTTGGTGACTGACTGGG
        TATAAGGAAGGAGGGGAACAGAGACCGGCAGCATGACTCCCAGCCTGCTGGGCTGGATGG
        CTGGTGGATGGTGAGTCCATTCACCAAACTGGGAGGCCCAGAGAGAGAAGCAGATTCTGG

24643   CTCTGTGGACATTCTCTCCTCCTTTGGCCCCCAGCTCAGCACAGTCTCCAGCTTTACTTC
        GGACTCAGACTATTCCTGCTCAGCCTTCGTTGCTGACTTCTCTGTTCTCCCTGAAACAGG
        AGTGTCTGCCCAGGCTCTGTCCTTGGCCTCTCCTCTTTTTACACTTCATTCTCTCCCTGG
        ACAATCTCTTCTCAGCCCAAAGCCCTAAATCTAAACCTTCAATTTCTGGTTGAAATCATT
        CTCCTGAGCTTCCAAAACTGTGGAGCACTGAAGAGGAGGAGATGGATGTGAGACATTTGG
        [A,G]
        TGACTTGGTGACTGACTGGGTATAAGGAAGGAGGGGAACAGAGACCGGCAGCATGACTCC
```

FIGURE 3P

```
        CAGCCTGCTGGGCTGGATGGCTGGTGGATGGTGAGTCCATTCACCAAACTGGGAGGCCCA
        GAGAGAGAAGCAGATTCTGGGCTATGGAGGATGAATGCAGGGTGGAGCATGTTGAGTCTG
        TTGTGCTCTTGGGACATCTGGATGGACATTTCCAGAAGGCATATGGGTATGTAAATCCAC
        ATAGTAGGCCAGCTGGCTGGAAATACAGATTTAGGAGACAGCAGAGTGAGGACGGGGATG

25329   TGAAGGAACTGCAAGCTGGGAGCGCTGAGATGACTGCCCTCCTGGTGCTTCCCAGGCCTC
        GGGGCTGTCCGTGTGGATGGGGAAGCAGATGGAGCCCTTGCACGCAGTGCCCCCGGCAGC
        CATCACCTTGATCTTGTCCTTGCTCGTTGCCGTGTTCACTGAGTGCACAAGCAACGTGGC
        CACCACCACCTTGTTCCTGCCCATCTTTGCCTCCATGGTAAGTAACCTGACAGTGGGGAG
        GAGCCCTTCCATTTCACAGGAACACATGGCCATATTGTGGGTCCCTGACGAGGCAGCAAT
        [G,A]
        TCCAGGCCAGACTCAGACCAGGCTTTGGAGACCCAGGTCTGACTGTGACGTGGATTTGTG
        GACCCTGGATGCCTCTGCCCCTGAGGCCTCCACTGCTTTGCCACTCCTCTTTGTACCCCT
        CCTGCTGACCAAAGCACCAACCATGGACCAAGTGCTCAAATTTATTTTATAAATCTAATT
        GGATTATTTTTCAAGCTGGGGAGACAGGACTTGGGCTAAGGAGGAGCAGGCCAGTGCCGT
        GGTCTCTGAGCATGTAGCACAGGTGTGCAGGAGGACTGCAGACTGGGAGCACCACTGGCT

25421   AGCCCTTGCACGCAGTGCCCCCGGCAGCCATCACCTTGATCTTGTCCTTGCTCGTTGCCG
        TGTTCACTGAGTGCACAAGCAACGTGGCCACCACCACCTTGTTCCTGCCCATCTTTGCCT
        CCATGGTAAGTAACCTGACAGTGGGGAGGAGCCCTTCCATTTCACAGGAACACATGGCCA
        TATTGTGGGTCCCTGACGAGGCAGCAATGTCCAGGCCAGACTCAGACCAGGCTTTGGAGA
        CCCAGGTCTGACTGTGACGTGGATTTGTGGACCCTGGATGCCTCTGCCCCTGAGGCCTCC
        [G,A]
        CTGCTTTGCCACTCCTCTTTGTACCCCTCCTGCTGACCAAAGCACCAACCATGGACCAAG
        TGCTCAAATTTATTTTATAAATCTAATTGGATTATTTTTCAAGCTGGGGAGACAGGACTT
        GGGCTAAGGAGGAGCAGGCCAGTGCCGTGGTCTCTGAGCATGTAGCACAGGTGTGCAGGA
        GGACTGCAGACTGGGAGCACCACTGGCTGGAAACCCCAGGAAGAGGCCTTGGAGGAGTGG
        GGACTTGGGAGTAGGTAGGAAGGGAGAGAGAATTCTGGGAAGATGGAGCAGCACAAGGAA

25797   TATAAATCTAATTGGATTATTTTTCAAGCTGGGGAGACAGGACTTGGGCTAAGGAGGAGC
        AGGCCAGTGCCGTGGTCTCTGAGCATGTAGCACAGGTGTGCAGGAGGACTGCAGACTGGG
        AGCACCACTGGCTGGAAACCCCAGGAAGAGGCCTTGGAGGAGTGGGGACTTGGGAGTAGG
        TAGGAAGGGAGAGAGAATTCTGGGAAGATGGAGCAGCACAAGGAAAGGCAATGGTGCACA
        TGACTGAGGACTCCTGGAAGCCTGGCTTGGTGAGCACAGGGATAAGGGATCCTGGGGAGT
        [T,G]
        GAGAGAGGTAGCTGTCGGTTGTGGGAAAAGCTGCTGAGTGCCAGGCTAAGGCATTCTGTT
        CTATGGACTAGCATGTTTTTTAGTTGGGAGTTAGAAGAAAGCAGAGCTTATAGGAAAATC
        AGTGGCTATGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATGCATTTCCTTCTGTC
        ATCCATTGCAAAGACGTACCAGCTTCAGGGTAGTATGGAAAGATCCCTGGTCTCGCAGTC
        AGAAGACCCGAGTTCAAGATGTGGGATCTCTGAACATGGCCCTTCAGTTCTTTCTTCCGA

25926   GGCTGGAAACCCCAGGAAGAGGCCTTGGAGGAGTGGGGACTTGGGAGTAGGTAGGAAGGG
        AGAGAGAATTCTGGGAAGATGGAGCAGCACAAGGAAAGGCAATGGTGCACATGACTGAGG
        ACTCCTGGAAGCCTGGCTTGGTGAGCACAGGGATAAGGGATCCTGGGGAGTGGAGAGAGG
        TAGCTGTCGGTTGTGGGAAAAGCTGCTGAGTGCCAGGCTAAGGCATTCTGTTCTATGGAC
        TAGCATGTTTTTTAGTTGGGAGTTAGAAGAAAGCAGAGCTTATAGGAAAATCAGTGGCTA
        [C,T]
        GGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATGCATTTCCTTCTGTCATCCATTGC
        AAAGACGTACCAGCTTCAGGGTAGTATGGAAAGATCCCTGGTCTCGCAGTCAGAAGACCC
        GAGTTCAAGATGTGGGATCTCTGAACATGGCCCTTCAGTTCTTTCTTCCGAGAGCTGTGC
        TGATGGCCAAGTAAGATGAGGGCTATGAAAAGCCTCTGTAGACTGCAAAATGAGCATGGG
        AGAGGCTGTCATTATTCTGGAATTGGGAGACAGATTTACAGAGGGCCTGAACACAGGATT

27289   AACAGGGGGTTTGCTGAAATAAATTATCATATATTCATATAATATGACATTATCAGGCCA
        TTAAAAATCACAGTTTCAAAGAGTAATAAAATGGGAACATGCTCATAGTATAGTTTTTTA
        AAATTGCAGATGGTATATGGCTAAAAATGTCTAATAATGCAAAGATGTATACAGACCTTA
        ATCCTCTAGCCTCCTCCCTAGAGATGACCTCTGTTAATTTCTCAAATATTTTTCTGGATA
        TTTTACACACTCACACACTTTTTTTGAGACAGAGTTTCACTCTTGTCACCCAGGCTGGAG
        [T,C]
        GCAATGGTGTGATCTTGGCTCACTGCAACCTCCACCTCCCGGGTTCAAGAGATTCTCCTG
        CCTCAGCCTCCCGAGTAGCTGGGATTACAGGTGCCTGCCACCTTGCCTGGCTAATTTTTT
        GTATTTTTAGTAGAGACGGGGTTTCACCACATTGGTCAGGCTGGTCTCAAACTCCTGACC
        TCAGGTGATCCGCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGC
        GCCCGGCCATTCATCTTAATTTTTAAAAAATCTAACCATGAAGCCTTGGTTATCTTGGAG

27591   CAATGGTGTGATCTTGGCTCACTGCAACCTCCACCTCCCGGGTTCAAGAGATTCTCCTGC
        CTCAGCCTCCCGAGTAGCTGGGATTACAGGTGCCTGCCACCTTGCCTGGCTAATTTTTTG
        TATTTTTAGTAGAGACGGGGTTTCACCACATTGGTCAGGCTGGTCTCAAACTCCTGACCT
        CAGGTGATCCGCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCG
        CCCGGCCATTCATCTTAATTTTTAAAAAATCTAACCATGAAGCCTTGGTTATCTTGGAGA
        [G,T]
        CTTTCCTGATTAGCACAAAAAGAAAAAAAAATCCAATTCTTTACAGCTGCATACTATTCC
        ATTATTTGTATGTGTCATATTTTATTTAACCATCCTGCTATTAGTGACCATTGAGTTGGC
        TTCCTGTGTTTTGCCGTTACATGGTTGCAACAAACATGTTTGCATGTGTCTGCCCTCATG
        TGCATGATACATGATTGATTTGATAGATTTTAGGAATTACATCATTCATTCATACACTCA
        GCAAATATTTAATGAGTGCCTACTCTCTGATAGGTGCTGTTGGATGTGGCTAAATTTTAA
```

FIGURE 3Q

```
28245   CATGTATAAAAACACACACATCCACAGATTTATATGCGGGAGAGAAGATGTGGTCCCTGG
        CCTCTAGGCTCTCTCAGTCTGTGGCAAGACAGACAGACATGTGCACGCGGCACTGTAAGG
        TTGAGCACAGTCTAAGTACTCAGCATGGTCTCTGGCACATAGTAGGTGCCCAAGAAATAC
        ATGTCGAATGAATTGAGGGGGTAAGGCCTTCTAGGGCAGGTGGCCTCTGACCTCAGCCTT
        CAGTGTTCCGTAGGTGGAATTATCTGCCAGAGACGTGGCAAAAGGGAGAGGAACCAAGAC
        [T,A]
        GAGGCACAGAGGTTCAAACGTACCCGGCACATTCAGAGAATCCTTTTCAGAATCACGTCC
        CCAAGAGCTTCTGTGTTCTGTACGGTGATGTTGCAGTGCTGTTTTTCCGCAGTCTCGCTC
        CATCGGCCTCAATCCGCTGTACATCATGCTGCCCTGTACCCTGAGTGCCTCCTTTGCCTT
        CATGTTGCCTGTGGCCACCCCTCCAAATGCCATCGTGTTCACCTATGGGCACCTCAAGGT
        TGCTGACATGGTAACACAGCTGTTTTTATTTACTCCCGTCGGACTATAACGCTGTTGTCA

29337   CAGCTCTGGGCTGGGAAAGAGGCAAGTGTTTGAGCCCAAGAGGCCAGAAATGTACCTGGG
        ACCAATCGGGTGTTCGTTATCTCAGAGCCTCTGCTGGGTATCTCAGGGACTCCATGAGCA
        TTTTCAAAAAAAAAGGTGGGTCCCAGAAACCATGGACTGCAAACTTGACTCCAATCCCCC
        AGTAAAATATCTACAACAGGGTAGTGAAGCGATGGTTAGTGACCATGAGGGAAGCTTGCA
        GAGCAGGCATCAGAAAGAGCCTGAGGAGGTCCACAGGGAAGCTGGCACGTCCTTGTAGGA
        [G,A,T]
        AGTTAAGGCACTGGGGTGAGCAATGAACCTGGACTCACGGAACACTGGGCTCTGTGACCG
        TTTCCCTGAATGGCCTAAGCTGTTGCCTCCTGTCACTTCTCTGAGGTCATTTTCCAAATG
        CGCACGGGCATAGAGAACCCATCCACTCTGCCTACTTCCCAGGGATGCCTTGAGCACTGA
        GGATACCTGGGGGACATGAAGTCGCACTGTCCTGGGGGTCGGGACACCCCAGCCAGGGAC
        AGAGCATGGCACAGGGACATCGAGGCCCAGTGAGCCGACCCTTTGTCCTCCTCTCTGAGA

29460   TCAAAAAAAAAGGTGGGTCCCAGAAACCATGGACTGCAAACTTGACTCCAATCCCCCAGT
        AAAATATCTACAACAGGGTAGTGAAGCGATGGTTAGTGACCATGAGGGAAGCTTGCAGAG
        CAGGCATCAGAAAGAGCCTGAGGAGGTCCACAGGGAAGCTGGCACGTCCTTGTAGGATAG
        TTAAGGCACTGGGGTGAGCAATGAACCTGGACTCACGGAACACTGGGCTCTGTGACCGTT
        TCCCTGAATGGCCTAAGCTGTTGCCTCCTGTCACTTCTCTGAGGTCATTTTCCAAATGCG
        [G,A,C]
        ACGGGCATAGAGAACCCATCCACTCTGCCTACTTCCCAGGGATGCCTTGAGCACTGAGGA
        TACCTGGGGGACATGAAGTCGCACTGTCCTGGGGGTCGGGACACCCCAGCCAGGGACAGA
        GCATGGCACAGGGACATCGAGGCCCAGTGAGCCGACCCTTTGTCCTCCTCTCTGAGAGCA
        CTAGTCCCCAGCAGGCCTCAGGGTGCTGACTCTGTCTCTTTTCCAGGTGAAAACAGGAGT
        CATAATGAACATAATTGGAGTCTTCTGTGTGTTTTTGGCTGTCAACACCTGGGGACGGGC

29994   CAGGAGTCATAATGAACATAATTGGAGTCTTCTGTGTGTTTTTGGCTGTCAACACCTGGG
        GACGGGCCATATTTGACTTGGATCATTTCCCTGACTGGGCTAATGTGACACATATTGAGA
        CTTAGGAAGAGCCACAAGACCACACACACAGCCCTTACCCTCCTCAGGACTACCGAACCT
        TCTGGCACACCTTGTACAGAGTTTTGGGGTTCACACCCCAAAATGACCCAACGATGTCCA
        CACACCACCAAAACCCAGCCAATGGGCCACCTCTTCCTCCAAGCCCAGATGCAGAGATGG
        [A,T]
        CATGGGCAGCTGGAGGGTAGGCTCAGAAATGAAGGGAACCCCTCAGTGGGCTGCTGGACC
        CATCTTTCCCAAGCCTTGCCATTATCTCTGTGAGGGAGGCCAGGTAGCCGAGGGATCAGG
        ATGCAGGCTGCTGTACCCGCTCTGCCTCAAGCATCCCCCACACAGGGCTCTGGTTTTCAC
        TCGCTTCGTCCTAGATAGTTTAAATGGGAATCGGATCCCCTGGTTGAGAGCTAAGACAAC
        CACCTACCAGTGCCCATGTCCCTTCCAGCTCACCTTGAGCAGCCTCAGATCATCTCTGTC

30207   CACCCCAAAATGACCCAACGATGTCCACACACCACCAAAACCCAGCCAATGGGCCACCTC
        TTCCTCCAAGCCCAGATGCAGAGATGGTCATGGGCAGCTGGAGGGTAGGCTCAGAAATGA
        AGGGAACCCCTCAGTGGGCTGCTGGACCCATCTTTCCCAAGCCTTGCCATTATCTCTGTG
        AGGGAGGCCAGGTAGCCGAGGGATCAGGATGCAGGCTGCTGTACCCGCTCTGCCTCAAGC
        ATCCCCCACACAGGGCTCTGGTTTTCACTCGCTTCGTCCTAGATAGTTTAAATGGGAATC
        [G,A]
        GATCCCCTGGTTGAGAGCTAAGACAACCACCTACCAGTGCCCATGTCCCTTCCAGCTCAC
        CTTGAGCAGCCTCAGATCATCTCTGTCACTCTGGAAGGGACACCCCAGCCAGGGACGGAA
        TGCCTGGTCTTGAGCAACCTCCCACTGCTGGAGTGCGAGTGGGAATCAGAGCCTCCTGAA
        GCCTCTGGGAACTCCTCCTGTGGCCACCACCAAAGGATGAGGAATCTGAGTTGCCAACTT
        CAGGACGACACCTGGCTTGCCACCCACAGTGCACCACAGGCCAACCTACGCCCTTCATCA

30497   AATGGGAATCGGATCCCCTGGTTGAGAGCTAAGACAACCACCTACCAGTGCCCATGTCCC
        TTCCAGCTCACCTTGAGCAGCCTCAGATCATCTCTGTCACTCTGGAAGGGACACCCCAGC
        CAGGGACGGAATGCCTGGTCTTGAGCAACCTCCCACTGCTGGAGTGCGAGTGGGAATCAG
        AGCCTCCTGAAGCCTCTGGGAACTCCTCCTGTGGCCACCACCAAAGGATGAGGAATCTGA
        GTTGCCAACTTCAGGACGACACCTGGCTTGCCACCCACAGTGCACCACAGGCCAACCTAC
        [T,G]
        CCCTTCATCACTTGGTTCTGTTTTAATCGACTGGCCCCCTGTCCCACCTCTCCAGTGAGC
        CTCCTTCAACTCCTTGGTCCCCTGTTGTCTGGGTCAACATTTGCCGAGACGCCTTGGCTG
        GCACCCTCTGGGGTCCCCCTTTTCTCCCAGGCAGGTCATCTTTTCTGGGAGATGCTTCCC
        CTGCCATCCCCAAATAGCTAGGATCACACTCCAAGTATGGGCAGTGATGGCGCTCTGGGG
        GCCACAGTGGGCTATCTAGGTCCTCCCTCACCTGAGGCCCAGAGTGGACACAGCTGTTAA

30738   TTGCCAACTTCAGGACGACACCTGGCTTGCCACCCACAGTGCACCACAGGCCAACCTACG
        CCCTTCATCACTTGGTTCTGTTTTAATCGACTGGCCCCCTGTCCCACCTCTCCAGTGAGC
        CTCCTTCAACTCCTTGGTCCCCTGTTGTCTGGGTCAACATTTGCCGAGACGCCTTGGCTG
```

FIGURE 3R

GCACCCTCTGGGGTCCCCCTTTTCTCCCAGGCAGGTCATCTTTTCTGGGAGATGCTTCCC
CTGCCATCCCCAAATAGCTAGGATCACACTCCAAGTATGGGCAGTGATGGCGCTCTGGGG
[G,A]
CCACAGTGGGCTATCTAGGTCCTCCCTCACCTGAGGCCCAGAGTGGACACAGCTGTTAAT
TTCCACTGGCTATGCCACTTCAGAGTCTTTCATGCCAGCGTTTGAGCTCCTCTGGGTAAA
ATCTTCCCTTTGTTGACTGGCCTTCACAGCCATGGCTGGTGACAACAGAGGATCGTTGAG
ATTGAGCAGCGCTTGGTGATCTCTCAGCAAACAACCCCTGCCCGTGGGCCAATCTACTTG
AAGTTACTCGGACAAAGACCCCAAAGTGGGGCAACAACTCCAGAGAGGCTGTGGGAATCT

30758 CCTGGCTTGCCACCCACAGTGCACCACAGGCCAACCTACGCCCTTCATCACTTGGTTCTG
TTTTAATCGACTGGCCCCCTGTCCCACCTCTCCAGTGAGCCTCCTTCAACTCCTTGGTCC
CCTGTTGTCTGGGTCAACATTTGCCGAGACGCCTTGGCTGGCACCCTCTGGGGTCCCCCT
TTTCTCCCAGGCAGGTCATCTTTTCTGGGAGATGCTTCCCCTGCCATCCCCAAATAGCTA
GGATCACACTCCAAGTATGGGCAGTGATGGCGCTCTGGGGGCCACAGTGGGCTATCTAGG
[T,C]
CCTCCCTCACCTGAGGCCCAGAGTGGACACAGCTGTTAATTTCCACTGGCTATGCCACTT
CAGAGTCTTTCATGCCAGCGTTTGAGCTCCTCTGGGTAAAATCTTCCCTTTGTTGACTGG
CCTTCACAGCCATGGCTGGTGACAACAGAGGATCGTTGAGATTGAGCAGCGCTTGGTGAT
CTCTCAGCAAACAACCCCTGCCCGTGGGCCAATCTACTTGAAGTTACTCGGACAAAGACC
CCAAAGTGGGGCAACAACTCCAGAGAGGCTGTGGGAATCTTCAGAAGCCCCCCTGTAAGA

31045 TGGGCTATCTAGGTCCTCCCTCACCTGAGGCCCAGAGTGGACACAGCTGTTAATTTCCAC
TGGCTATGCCACTTCAGAGTCTTTCATGCCAGCGTTTGAGCTCCTCTGGGTAAAATCTTC
CCTTTGTTGACTGGCCTTCACAGCCATGGCTGGTGACAACAGAGGATCGTTGAGATTGAG
CAGCGCTTGGTGATCTCTCAGCAAACAACCCCTGCCCGTGGGCCAATCTACTTGAAGTTA
CTCGGACAAAGACCCCAAAGTGGGGCAACAACTCCAGAGAGGCTGTGGGAATCTTCAGAA
[G,-]
CCCCCCTGTAAGAGACAGACATGAGAGACAAGCATCTTCTTTCCCCCGCAAGTCCATTTT
ATTTCCTTCTTGTGCTGCTCTGGAAGAGAGGCAGTAGCAAAGAGATGAGCTCCTGGATGG
CATTTTCCAGGGCAGGAGAAAGTATGAGAGCCTCAGGAAACCCCATCAAGGACCGAGTAT
GTGTCTGGTTCCTTGGGTGGGACGATTCCTGACCACACTGTCCAGCTCTTGCTCTCATTA
AATGCTCTGTCTCCCGCGGAAAGCTCCACTGTGCTGCTGACTTGTCTCTGGTTTTCTGCA

32591 AGTGTATGGCCTCCCCAACCAAACCCTTTGCTTCTTAGGGCAAGGACCACCCTGTCTCAT
TGATCACTGTCCTGAGCCTATCTCAGGGGAGGCAAAAGAGAGGGACCTGTATTCAGAGAT
CTTCCCTTGGCATGACTTGCTTTTGGCCACTTACCTTTCCCTACAAGCTCTATGAGGCCA
AGGCCCTTCATGGTTAGTGTAAGGAGCAGTGGGCATGGAGTTGGAAGATCTGGGTTGGAA
CGGTAACTGCCACTAACTCGATGTGTGATTCTGAACACTTAACTTAGCCATACATGCTCT
[C,T]
TTATTTGCTTTTGATGGCAAATAAGAGAAGGCCCAGCAAACAGTGGCTTAAACCAGAAGG
TCAATTAATGTTTACTTTTCAGGAAGTCTGTAGGTAGATGGTTGCTGGCATTGGCCCAAC
AGCTCATTTCAGCCTCCAAGGACTTGCGCTCCATAGTCCACTCTGTCATCTTAAAGCCTT
CACACTTTTACCCCCATGCTTGACCCCCAGGCTACATACACAGCT

Bac accession number: AC034305
Chromosome: 17

FIGURE 3S

SEQUENCE LISTING

<110> PE CORPORATION (NY)

<120> ISOLATED HUMAN TRANSPORTER PROTEINS,
NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
AND USES THEREOF

<130> CL000662PCT

<140> TO BE ASSIGNED
<141> 2001-12-05

<150> 09/729,094
<151> 2000-12-05

<150> 60/211,220
<151> 2000-06-13

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 2223
<212> DNA
<213> Homo sapiens

<400> 1
gtctccctcc cgcgcgatgg cctcggcgct gagctatgtc tccaagttca agtccttcgt 60
gatcttgttc gtcaccccgc tcctgctgct gccactcgtc attctgatgc ccgccaagtt 120
tgtcaggtgt gcctacgtca tcatcctcat ggccatttac tggtgcacag aagtcatccc 180
tctggctgtc acctctctca tgcctgtctt gcttttccca ctcttccaga ttctggactc 240
caggcaggtg tgtgtccagt acatgaagga caccaacatg ctgttcctgg gcggcctcat 300
cgtggccgtg gctgtggagc gctggaacct gcacaagagg atcgccctgc gcacgctcct 360
ctgggtgggg gccaagcctg cacggctgat gctgggcttc atgggcgtca cagccctcct 420
gtccatgtgg atcagtaaca tggcaaccac ggccatgatg gtgcccatcg tggaggccat 480
attgcagcag atggaagcca caagcgcagc caccgaggcc ggcctggagc tggtggacaa 540
gggcaaggcc aaggagctgc cagggagtca agtgattttt gaaggcccca ctctggggca 600
gcaggaagac caagagcgga agaggttgtg taaggccatg accctgtgca tctgctacgc 660
ggccagcatc gggggcaccg ccaccctgac cgggacggga cccaacgtgg tgctcctggg 720
ccagatgaac gagttgtttc ctgacagcaa ggacctcgtg aactttgctt cctggtttgc 780
atttgccttt cccaacatgc tggtgatgct gctgttcgcc tggctgtggc tccagtttgt 840
ttacatgaga ttcaatttta aaaagtcctg gggctgcggg ctagagagca agaaaaacga 900
gaaggctgcc ctcaaggtgc tgcaggagga gtaccggaag ctggggccct tgtccttcgc 960
ggagatcaac gtgctgatct gcttcttcct gctggtcatc ctgtggttct cccgagaccc 1020
cggcttcatg cccggctggc tgactgttgc ctgggtggag ggtgagacaa agtatgtctc 1080
cgatgccact gtggccatct ttgtggccac cctgctattc attgtgcttt cacagaagcc 1140
caagtttaac ttccgcagcc agactgagga agaaaggaaa actccatttt atccccctcc 1200
cctgctggat tggaaggtaa cccaggagaa agtgccctgg ggcatcgtgc tgctactagg 1260
gggcggattt gctctggcta aaggatccga ggcctcgggg ctgtccgtgt ggatggggaa 1320
gcagatggag cccttgcacg cagtgccccc ggcagccatc accttgatct tgtccttgct 1380
cgttgccgtg ttcactgagt gcacaagcaa cgtggccacc accaccttgt tcctgcccat 1440
ctttgcctcc atgtctcgct ccatcggcct caatccgctg tacatcatgc tgccctgtac 1500
cctgagtgcc tcctttgcct tcatgttgcc tgtggccacc cctccaaatg ccatcgtgtt 1560
cacctatggg cacctcaagg ttgctgacat ggtgaaaaca ggagtcataa tgaacataat 1620
tggagtcttc tgtgtgtttt tggctgtcaa cacctgggga cgggccatat ttgacttgga 1680
tcatttccct gactgggcta atgtgacaca tattgagact taggaagagc cacaagacca 1740
cacacacagc ccttaccctc ctcaggacta ccgaaccttc tggcacacct tgtacagagt 1800
tttggggttc acaccccaaa atgacccaac gatgtccaca caccaccaaa acccagccaa 1860
tgggccacct cttcctccaa gcccagatgc agagatggtc atgggcagct ggagggtagg 1920
ctcagaaatg aagggaaccc ctcagtgggc tgctggaccc atctttccca agccttgcca 1980
ttatctctgt gagggaggcc aggtagccga gggatcagga tgcaggctgc tgtacccgct 2040
ctgcctcaag catcccccac acagggctct ggttttcact cgcttcgtcc tagatagttt 2100
aaatgggaat cagatcccct ggttgagagc taagacaacc acctaccagt gcccatgtcc 2160
cttccagctc accttgagca gcctcagatc atctctgtca ctctggaagg gacaccccag 2220
cca 2223

<210> 2
<211> 568
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Ala Ser Ala Leu Ser Tyr Val Ser Lys Phe Lys Ser Phe Val Ile
 1               5                  10                  15
Leu Phe Val Thr Pro Leu Leu Leu Leu Pro Leu Val Ile Leu Met Pro
            20                  25                  30
Ala Lys Phe Val Arg Cys Ala Tyr Val Ile Ile Leu Met Ala Ile Tyr
        35                  40                  45
Trp Cys Thr Glu Val Ile Pro Leu Ala Val Thr Ser Leu Met Pro Val
    50                  55                  60
Leu Leu Phe Pro Leu Phe Gln Ile Leu Asp Ser Arg Gln Val Cys Val
65                  70                  75                  80
Gln Tyr Met Lys Asp Thr Asn Met Leu Phe Leu Gly Gly Leu Ile Val
                85                  90                  95
Ala Val Ala Val Glu Arg Trp Asn Leu His Lys Arg Ile Ala Leu Arg
            100                 105                 110
Thr Leu Leu Trp Val Gly Ala Lys Pro Ala Arg Leu Met Leu Gly Phe
        115                 120                 125
Met Gly Val Thr Ala Leu Leu Ser Met Trp Ile Ser Asn Met Ala Thr
    130                 135                 140
Thr Ala Met Met Val Pro Ile Val Glu Ala Ile Leu Gln Gln Met Glu
145                 150                 155                 160
Ala Thr Ser Ala Ala Thr Glu Ala Gly Leu Glu Leu Val Asp Lys Gly
                165                 170                 175
Lys Ala Lys Glu Leu Pro Gly Ser Gln Val Ile Phe Glu Gly Pro Thr
            180                 185                 190
Leu Gly Gln Gln Glu Asp Gln Glu Arg Lys Arg Leu Cys Lys Ala Met
        195                 200                 205
Thr Leu Cys Ile Cys Tyr Ala Ala Ser Ile Gly Gly Thr Ala Thr Leu
    210                 215                 220
Thr Gly Thr Gly Pro Asn Val Val Leu Leu Gly Gln Met Asn Glu Leu
225                 230                 235                 240
Phe Pro Asp Ser Lys Asp Leu Val Asn Phe Ala Ser Trp Phe Ala Phe
                245                 250                 255
Ala Phe Pro Asn Met Leu Val Met Leu Leu Phe Ala Trp Leu Trp Leu
            260                 265                 270
Gln Phe Val Tyr Met Arg Phe Asn Phe Lys Lys Ser Trp Gly Cys Gly
        275                 280                 285
Leu Glu Ser Lys Lys Asn Glu Lys Ala Ala Leu Lys Val Leu Gln Glu
    290                 295                 300
Glu Tyr Arg Lys Leu Gly Pro Leu Ser Phe Ala Glu Ile Asn Val Leu
305                 310                 315                 320
Ile Cys Phe Phe Leu Leu Val Ile Leu Trp Phe Ser Arg Asp Pro Gly
                325                 330                 335
Phe Met Pro Gly Trp Leu Thr Val Ala Trp Val Glu Gly Glu Thr Lys
            340                 345                 350
Tyr Val Ser Asp Ala Thr Val Ala Ile Phe Val Ala Thr Leu Leu Phe
        355                 360                 365
Ile Val Leu Ser Gln Lys Pro Lys Phe Asn Phe Arg Ser Gln Thr Glu
    370                 375                 380
Glu Glu Arg Lys Thr Pro Phe Tyr Pro Pro Pro Leu Leu Asp Trp Lys
385                 390                 395                 400
Val Thr Gln Glu Lys Val Pro Trp Gly Ile Val Leu Leu Leu Gly Gly
                405                 410                 415
Gly Phe Ala Leu Ala Lys Gly Ser Glu Ala Ser Gly Leu Ser Val Trp
            420                 425                 430
Met Gly Lys Gln Met Glu Pro Leu His Ala Val Pro Pro Ala Ala Ile
        435                 440                 445
Thr Leu Ile Leu Ser Leu Leu Val Ala Val Phe Thr Glu Cys Thr Ser
    450                 455                 460
Asn Val Ala Thr Thr Thr Leu Phe Leu Pro Ile Phe Ala Ser Met Ser
465                 470                 475                 480
Arg Ser Ile Gly Leu Asn Pro Leu Tyr Ile Met Leu Pro Cys Thr Leu
```

```
                485                 490                 495
Ser Ala Ser Phe Ala Phe Met Leu Pro Val Ala Thr Pro Pro Asn Ala
            500                 505                 510
Ile Val Phe Thr Tyr Gly His Leu Lys Val Ala Asp Met Val Lys Thr
        515                 520                 525
Gly Val Ile Met Asn Ile Ile Gly Val Phe Cys Val Phe Leu Ala Val
    530                 535                 540
Asn Thr Trp Gly Arg Ala Ile Phe Asp Leu Asp His Phe Pro Asp Trp
545                 550                 555                 560
Ala Asn Val Thr His Ile Glu Thr
                565
```

<210> 3
<211> 32816
<212> DNA
<213> Homo sapiens

<400> 3

```
ttcaaccatt gtggaagaca ctgtggcgat tcctcaagga tctagaacca gaaatatcat 60
ttgacccagc aattttatta ctgggtatat acccaaagga ttataaatca tgctgctata 120
aagacacatg cacactattt acaatagcaa agacttaaaa ccaacccaaa tgtccatcaa 180
tgatagactg gataaagaaa atgtggcaca tacataccat ggaatactat gcagccatta 240
aaaataatga ggtcatgtcc tttgcaggga catggatgaa gctggaagcc atcattctca 300
gcaaactaac acaggaacag aaaaccaaac accacatgtt ctcagtcata agtgggagtt 360
gaacagtgag aacgcattga cacagggagg ggaacatcac acacgggggc ctgtcagggg 420
gttggagggc aaggggaggg agagcattag gacaaatacc taatgcatgt gggtcttaaa 480
acctaaatgt ccggttgata gctgcagcaa accaccatgg cacatgtata cctatgtaac 540
aaacctgcac attctgcaca tgtatcccag aacttaaagt aaaattaaaa aaaaagaaaa 600
gaaaaaagaa ctgaagttgt ttacttgctc tcattcatgc atcccggaga aaaaggtttg 660
agtgcacatc ctggattagg cactgagaaa ggcactagct ggacaggtgg tgatgaataa 720
aacagacagt aaatagaaat tacatcataa taatgtgtca tatattttaa aatagctaca 780
agatatttta aatgttctca ccacaaagaa atgacaaata tttgggccag acgcggtggc 840
tcacgcctgt aatcccagca ctttgggaga ccgaggtggg cggatcacct gaggtcagga 900
gttcgagacc agcctggcta acatggtgaa accccatttc tactaaaaat gcaaaaaatt 960
agccgggcgt ggtggtgcac acctgtaatc ccagctactt gggaggctga agcaggagat 1020
ttgcttgaac ctaggtggca gaggttgcag tgagccgaga tcgtgccact gcactccagc 1080
ctgggtgaca ggagcacaac tctgtctcaa acaaacaaac aaaaaacaaa aacaagagaa 1140
atgataaata ttcgagtgat aaatatgctc attagcctga tttgaacaca ccacaattat 1200
acacacattg aaaaatcaca tggtaccccg taaatataga caatgatttg tcaattaaaa 1260
atgaaataac acttaaaaaa taaaaaagta aaaagtaaaa attacaccaa taaatataag 1320
aggtacaaat tgtgctaagt gccctgggga cacaggaaag gcgggaaaac ccagggctat 1380
atgcatgaga gttacaaagg gaaaaggaca ggagggaggc aattgcagga ggggcttggg 1440
agaatgcatg tccttgggtg caggtcacag gaaggaactc atgagcttga ttcaggatgt 1500
gttgaatttt cgggccgaga cacgtccagt ctgcggaagg ctggacatct gggactctgg 1560
catcatggct gggttgaagg cagaggatgg taatcactaa ggagccggct gtggttaggc 1620
caccagcatg gatgagactc cccaaaagga agctgcagaa tgagaggcag gcagaggaga 1680
ggaaagaaga aaatcacaga ggtggggatg tctttgcatc cgtgtgtctc cagtgcccaa 1740
aacagggcct cgcggagaag aggtgctcgg cacctgtctg ttgcctggcg ggctgaatga 1800
atacatgggc gactgtctca gtgtcgcctt agttgtgtcc cttcctctct agagctccgt 1860
ttccctctga cctgggtcgg gcgggcagct gcggctgctg aggctcggtg gggcccctcc 1920
aagacgcgtg tccgcatctg cccgccgggc gtctgcgggg tgcagcgtcc actggagcgc 1980
gacagcccct gggacagagg aggacagtgg cctcgcttcc ctgtgcgatc gcccaggagc 2040
tccgggccgg agagtgcgag cggggaaaag gggtcctgca cctagagtgg ggcggacgtg 2100
gcgaggaagc caggggggac cgggaagcga ggcccgcggt gcggagggcg cggggcgtgg 2160
ggggacacct ctcggagaga caccggaggg gcggaagtaa ggagatggaa aggagaggga 2220
gatcggggag atagacctga gagacccaga ggcctgcaga gagtttcatc cgggaccctt 2280
cagagcccag gaaagagcag atgcggacgc gggagggcgc cttacgccaa agcgggcagc 2340
accagtgacc aaaacacgcc ccgcttggca gccccgggac gcacctctgc ctcggcagcg 2400
caggagaggc ttggacagcg cgagatgcta gggcccaggc tgcccctaga gggctggccc 2460
gaagcgttgg agtccaaaga cgcctcccac cgccgccggg tggcagaatt gggggcaggc 2520
gcgtcccaca gaccccgagg ggtggccccg ccccagggcc gcggggaggc gcccccgtgc 2580
ggggcggagt tgtcaccgcc ccctccccaa tccccgggga ctgtggcccc ttcttaagcc 2640
cgcggcgcct ctagctgccc ctcactcgtc tcgcccgcca gtctccctcc cgcgcgatgg 2700
cctcggcgct gagctatgtc tccaagttca agtccttcgt gatcttgttc gtcaccccgc 2760
tcctgctgct gccactcgtc attctgatgc ccgccaaggt cagttgcatc tccaggcagc 2820
ccttcggaca cccggcgtcc tgtgcccact aacgggcacc gatcccggga gccctgagct 2880
ggagcgcacg gatttcgcgg ggagcacagc tctcccgggg cgcgcgcact caggagctcc 2940
```

aggtgccgga tgggaggtgc cctgtaaaga atctgagggg catggcgacc ccagggcgca 3000
ccacccttgg ggtttacaga tcccagggcg caagagccgt ccaggcaagc acggaaacct 3060
cgaagtgagc acagatctca gccacacaga tcccagcctt aggctcagcc cctggctccg 3120
aatcgaatct cccacagtgc ataaccctgt ttcccccaa aatgccacct gcgccaacag 3180
ggaacctggg agcttgcctt tccctctctc tctcctgtct tttcccttcg ccaaagaaga 3240
cttcaagctg taggtggctt ctgccgtcag gagggaccta caggaaaaaa atcatcaccc 3300
acgtggatcc tgcgctgtct ttgccactct ctggcccttc cttgggcctt agtgtctcta 3360
tctatgatcc acattccttc caacctggag agccacatct gattcataat cctgctccag 3420
gctgcaggca ggttggggtt ggcctgcttt gcctcctgcc tgctggggct gtagcaggag 3480
gcgggacaca ttcccagagc tcgcagcctt gggtggcagg acctggagtt gcagggaagc 3540
ttcctcccag gccctagtct cctaatgctt ctgtgaggga gagagagaat gaatggcctt 3600
ggccgcaggg tgggcgcagg ctccactggg ctgtgcacag ccagtttggc ggaggcccaa 3660
gccctttgaa gccttttgtg gctgctggct gctccttcct cgttccttct ttcagccctt 3720
tcactctcag cccagacagg aaacctccag ctccccacct cccctcccca ggcaggtttg 3780
ggaaacagag gagctcttca ggggatgctc tgggggggag ctctagagga agggagtgca 3840
ctggggtgtc aggaagccaa cctgcaagag aactggactt ccaccattct gtcatctgtg 3900
tgactttagc caagtgcctg tgctctctgg gccttggtgt tctcatctgt acaatggggc 3960
taggtggttt ggttctgaca ctttaaggct ttgggaatca aaagggatta acgttagttc 4020
ctaggatggg ggaggggggag actgggagga gcccttgggt gggcctagca caggccctgg 4080
atgggtcaag gacagcagat ccaactgtgg aggctgtgca gttgccacac caactgtggg 4140
cagccacact tcctctgtag caagaagtct ggggttgtta ttgtccaggg gaagtagcca 4200
ggcaggagag ctgcgatttc agctctgctg gtagggagtg atgttccctg gaatgattat 4260
agtagcttgg ctgaccttcc tgccacagga gaccccactc acacacacac acatacacac 4320
tctctctctc acacacacac actcacaaac acacacagac acacacacaa acacacagac 4380
acacacaaaa cacacgcaca aacacaaaca cacacaaaaa cacacgcaca aacacacacc 4440
caaacacaca aacacagaca cacaaacaca cacacaaaca cacacacaaa acacacaaaa 4500
aacacaaaca cacacataca catacacaca cacacacaca cacaccctga actggaaacc 4560
ctaactcagt gtgtgtgtat gtgtgtgtat gtgtgtgtgt gtaagagaga gagagagaga 4620
gattaagctg tcctttgagt gaggaccagg gaggggaaga agagaaccca gggagagtcc 4680
ttccaaaggc tgccttcacg agctttcctt ctggcggggt tgggtgagga ccctggacct 4740
tgtcttcttg ttttttccct ttctgcctgt tttggtcacc ctgcccccac cctccatggc 4800
cgccccattg tgcaaggaaa cccagagggt acacagcacg ggcagggcag ctgggaagct 4860
ggtgagaagc tgggaggacc ttggcagcct gagcaacaca gtccttgcca ggaggtgact 4920
cccagggcac gccacccctct gccaacaccc aggcctctct cctcaccgac tgtctccagt 4980
tttcctgtct ccacctggat tccctcctgg cctcatctct gctccactct ctctatcctt 5040
cctctgggtc ttttttaat tgaaaaaaaa tttaatgaaa taaatgatag atttcttgta 5100
tcacttattt tattaaaatg taaaaggttt ctttttgca aatctgtaag atataaagta 5160
aaaataaaag tacactcaaa tcccataagt tattcacatt ttgatgaaca tctttccaga 5220
tgaatctctc tctctcttcc cagacacaca cacacacaca cacacacagt aggttttgcc 5280
tgcatttttt cattaagtgg tgtgtcagga caccctgcct tgttaatgtt gaacttttct 5340
aacatccgct tcccatcctg ccctctccct tgacactgtg gaggcattct agactagggg 5400
ggtcaagcct gttgacttca gggatgaggc acctcctggg cttctaaata gtggcgcgga 5460
ggtgaggggg cagttaacct tgtgtctcgt cctctttcct agtgggtctg cttgactcct 5520
ccaggaacgc acagtgtaca ttggtgacgc acgccaccta ctgcttctaa gtttagagaa 5580
tcaaaagtta ccgaggactt tgtgcgccat atgggaagaa tgagcactct taaatccacg 5640
atttgcagat gaagacatga aacaagaggg gacagggacc aggattggga gcaggaggag 5700
taatttatga gcgacattgt ttagaattgc tatcacttga tgatagtaag aagcaaacta 5760
atttttagct aatattattg ttttaaaatt ctctccaatg cgccctctca ttgtctgccc 5820
ctggaggcat cattctgatg gcctgcccag ggtacacccc gacaccaagc cccaaggaag 5880
ttagtggctg ccaaaggcca gacagtggct gacagtgggc gccaatcata tctgtctggt 5940
gtcaaagcct gggcttccag tcacgctgct gttccgcctt tagttcagcg gctgtcaacc 6000
agcattgacc ctttctcctg ccctcaccct gccccacaac aggggaactt tggcaaagta 6060
cagagacatt ttttgctgtc caacctggag acagtcttac tggcatctca taggtggagg 6120
ccagcggtgc tctaaacacc ctgcagtgca cagctcccac aacaaagcat catttagccc 6180
aaaatgtcag tgtgccgagg ctgagggacc ctgcctccca gtagggaggt gccctggttt 6240
gctcgtggga tgctgaaaaa agatttattt ttttgtggct gataacacaa ccctgacaaa 6300
gaatttccaa gtcttcctgc actgttttgt gcaaataata catacgctct tctgggtgat 6360
gagaagcagg gattgtgtac aggtgcatct gttcttcagc agcattgtca gagttaaact 6420
cagatgaatg ctattgattc ctttaataaa catttgcaaa gatggccggg cacagtggct 6480
catgcctgta atcccagcac tttgggaggc cgagacaggt ggatcacgag gccaggagat 6540
caagaccatc ctggccaaca tggtgaaatc ccctctctac taaaaataca aaaattagcc 6600
gggcgtggtg gcgcttgtct gtagtcccag ctactcagga gactgaggca ggagaatcgc 6660
ttgaacctgg gaggtagagg ctgcagtgag ccaagattgc accactgcac tccagcctgg 6720
ggacagagca aggctctgtc tcaaaaataa gtaagtaagt aaataaataa ataaataaat 6780
aaataagcaa gaatttgcaa agatatccta agtgttgggc ctgttctgga tgctgaggac 6840
ggtgatctac aaatacagca ggttcttgaa taatgttgat tcattcaata tcatttcatt 6900
ataatgttga tgagggaggg aaaaaaaagg aaggatccct tgagcccagg agatggaggt 6960
tacagtgagc tgtgaccgtg ccacttcact cccacctggg caacagagcc agaccctgtc 7020

tcaaaaaaaa aaaaaaagaa ataaaagagc gagagagaaa gaaaagaaaa tgattactgg 7080
ctggggccac tgtctgtgtg gagcgtgcac attaccctcg tgtccacatg gcttttcttt 7140
ggctagtatg gtttccttcc acatcccaaa cccgtgcacg ttaggtgaat tggagtgtct 7200
gtatggtccc tgtctgagtg agcgtgggcg tgcgtgtcag tgtgcattct gcaatgggat 7260
ggcatcttgt ccagggctgg tttccacctt gtaccctgag ctgccgggac aggatctggt 7320
cacccaagac cctgacctgc tgtaactggg taaataatta tctaacttgt tttcaatgtt 7380
tcttaagtat atgtatagct cacattccct tcagtgttta atattggaag tgttttggtc 7440
tttatttaag atctccgtga tgtttttgtg accagaaata tgctgtagaa atttaactgt 7500
tgtttctagc aatttgccta tgggaatatt ggcttatgtt gtttcgctta cgcattgcaa 7560
tttccaaaaa ccaatcaatg atgttaagtg aggactcact gtactgtttg tgctttcgag 7620
tcacgcactg gttgtggtgg tagaaggaca gttgaggaaa cagtgacaac tccatatgct 7680
aatggctggg gagggtactc agaggaaggg cacaaaccag actatagaag aggcgcaggg 7740
agacatctaa gaaggaactc tgaggttggg cgcggtggct cacgcctgta atcccagcac 7800
tttagggtgc tgaggtgggc ggatcatgag gtcaggagtt cgagatcagc ctggccaatg 7860
tggcaaaacc ccgtctctac taaaaataca aaaattagca gggcgtggtg gcaggtgcct 7920
gcaatcccaa ctccggaggt tgaggcagga gaatcgcttg aacccgggag gtggaggttg 7980
cagtgagctg agattgtgcc attgcactcc agcctgggca acaggatcga aactctgtca 8040
cacacacaca cacaaaaaat actgatgaaa cataaaacaa cctagggagg tggctagttt 8100
tatcacataa ttattattac ttttatttca atagctttag gggtacacgt agttttcggt 8160
tacatgaatg aattggatag tggtgaagtc tgagatttta atccctccct cctatcccac 8220
cctgtctgct tctaagtctc cagtatccat tcgaccacgc tatatacctc tggataccca 8280
tagcttagct cccacttata agggagaaca tgcactattt ggctttccat tgctgagtca 8340
cttctcttag aacaatggcc tcctaggcgg caagagcgac actccatctc aaaaataaaa 8400
taataataaa accaaaaaaa ccaggtattt tattcttctt ctccttctcc tcttcctcct 8460
ttcttttctt ccttctccat ccccttcct ctctttcttc tatcccctcc tcctccttct 8520
cctccttctc cttctttctc cttttcttcc ttgtcctatt cttgatcttt tcttttgaga 8580
ggcagctaat ccaaggtttg agaagatgaa agaacgtgcc tagaaccaca cagctgggaa 8640
ggagggaggc agggaggagg ggtgggaatg gggcaggagt cctttgcgaa tagatccctg 8700
gcctgacccg ggaaagctgt gctgaccagg gctggggaac aagatgactt tgaggggaat 8760
ccctctgaga tcagcactgt gtcttgacaa tccatgccag ccgccgtccg gagtgttctg 8820
ggggtgggga gagggaggcg gcaacacgct gaggcctcag gactgtctct tcagtttgtc 8880
aggtgtgcct acgtcatcat cctcatggcc atttactggt gcacagaagt catccctctg 8940
gctgtcacct ctctcatgcc tgtcttgctt ttcccactct tccagattct ggactccagg 9000
caggtgagca gacccaaggg atcctggtga ctttctggtt ctccccttct ctctttctct 9060
agtccccact gtgagtcgca caggcctggg ggtgaccgga aaaccctcat ttgtggattc 9120
tccctggcag ggagacacca ctcgagcctg catccccact ccaagctgtc cctgaagtca 9180
gcatctgggg actgggtggc tctagtgtgt ggcaagggac agtcctgatg aggccttcgt 9240
gccacgctcc aggtgtgtgt ccagtacatg aaggacacca acatgctgtt cctgggcggc 9300
ctcatcgtgg ccgtggctgt ggagcgctgg aacctgcaca agaggatcgc cctgcgcacg 9360
ctcctctggg tgggggccaa gcctgcacgg taattacgcc ttctctctct tgccacgtgg 9420
ctctgcatga gccccagggc tggaaggggg tggaggatgg cacagaccag gccatccact 9480
ggtgagggct ggccatgggc ttacctggac ttggctgggt ggggtgcagt tatagcttta 9540
gtgggagaga ccagatgcat gcgtggtggt ggcacatggt gagcagcagt aagtaagggt 9600
cctcgaatcc agaggaggtg ggtcagcaag agtccttgca ggcttggaag gctttctggg 9660
ggaggcagct agctgcaggg ttccaccggg aacaaattgg atagaggctg gatcaagctg 9720
tgtctgatag gataagggaa gcaggccaga agtggctcaa ctacccagct catggggaag 9780
cagaaaggtc ctctctccaa gctggagcat ctattcccac tgcaaagaag cttcttatct 9840
tccccgatat cactcagtac cccagcttct ctctccattt ccaggatctc tcctgccaat 9900
ctagctagcc atttccagct aagccatgga gtcaatataa tcataatcat aaccataatc 9960
aatcatgatc ataatgggta tattgagtgt ctacaaggcc ccaggcatga taccaggagc
10020
ttatgaattg cctcatttaa ttcttaccac aactctgaga gctgagtatt cttactgccc
10080
acattctgtg gatgagggat tggaggcaga gagggataaa gtgatttgct catggacaca
10140
cggggattgg acccagcttc tctgataagg cctgtgtcct ctctaatcag aaactcaggg
10200
catatcttcc ttttgagaca atgtgtcccc tcaatgatgg cacgtccttg gcccagccat
10260
caggagtcag ctgctggtca gtttacggta aattcctcct gaggcgcccc tgtgtcaggg
10320
gctgtgctag accctgagca cacacagacg ctagcctgcc ctgcaggcaa cccatgcccg
10380
gggctccctt ggtgcctcag attcccctga gcaggggact ttgtggctct ctgatctgtt
10440
acacatcctg gcattgattc cttccagtaa ttgctttagc atcaaatcaa aagccatcat
10500

attttctaga aatgagagac cccaggaaag tggacctcag ggccctcaga attcttctgc
10560
ttggctccct tgagtggcca gcttgggtgg gaggccactc cagtgggttt cattctgcag
10620
catgctggag agcttccact tccaaaccca agttcacaca tgcttctgta tccttcctgc
10680
caccttgctc ctctgagtat ggtctccggt tgtccaaggc actgcctgtc ctgggagtca
10740
cctgtatgtg aggcaccctt ggtgccttga gatatcatgt agaagccttg gttcttctca
10800
gacaactcca ttcatgcaaa ctctcccccc cctcctagcc tgggtcccgg gctttgtttt
10860
tttttgggtc cataatgtct gcctgtgtgg acagcagctt gggccctggt gcagaacagc
10920
tcctaggtcc cttcttcagg ctcctacccc tgcccctgct cctaccccca ggtgaattag
10980
gagccctgag gaggagcctg gctgcagcga ggcccacaga ctgagagtag ctgagctcct
11040
tctgtcccta gccttggaca gctggggcat gtagagccac agagcagagt caggccctgc
11100
cctgctcaca gcccaaggag agagcagaca tggaaacagg tgctttgaac ccagcacagc
11160
gatgattaga gtagggggaa ggattgagaa gggtcaggcc agccccacct ggtgcacaca
11220
ctgagagcgt ggtcccagag gagggatgtt gtttgagcag gctctgaagg accatgagga
11280
gtcttcctga tagacagcag agaagggagc aggggttaca agcaaaggga gtgtttcttc
11340
tgaatactgt ttgtgtgata gctccactgc agcatggagg ggtcaaagtg tatgtgcggg
11400
gcggagggga gatggcaggg ttggagtggc agccgggaga atagtcacac tttcccaagc
11460
tccctcccca gctcacccta cccctactct gcttagccct tctgaacttc tgagaggtgc
11520
aacagagttt gggggtgggt gggaatttcc tagccagaag tgggaagctg gggctgcctg
11580
cacatagggg tattccagca caccctaggg caagctcata ttgagttggc accatctgga
11640
tgcctgggct tcccctgcta gatggtgggg caggggtgct ccttagaacc acgactggat
11700
ctgaggcctc ttggtaaccc cagaagcaag cagagtagac atcagtcatg ggtgtgggag
11760
aggcaggagg gagagaggaa tggaggaagc aaagaaggga aggagggagg gaggggaggc
11820
tctaaaaccg tcatccctat tccaatatct gatcttgaat tggcctcaac acctgtgcat
11880
ccctgcaggg gtggacccag tccccagttg cttcccaggg agtacggggg tggggtgggg
11940
attctctggc tttcctccct gcccctcctc tgcaggctga tgctgggctt catgggcgtc
12000
acagccctcc tgtccatgtg gatcagtaac acggcaacca cggccatgat ggtgcccatc
12060
gtggaggcca tattgcagca gatggaagcc acaagcgcag ccaccgaggc cggcctggag
12120
ctggtggaca agggcaaggc caaggagctg ccaggtgagc ccctggccag ggcactgcca
12180
ggccacaaca gcagccttcc cctccctctg ctggcaaatg ctttggccac ctccttctcc
12240
ctgtctgctt cccggagccc tcctttaaac acgcatagag aaaaaaaaat agaaaatact
12300
gttgtcctaa gttttaggag gggattattg cacacaactt agatccttta atagagcttt
12360
gaacaaagtc tcaccctcag ttcccatcag ttgcagaaat cagtgtgttc acctgattat
12420
tcatttgggc atctttcgag cacttaggga tgcccctcac tccttgctac tcctgctcat
12480
cctcaaggag gccttttctg acctcctcga gcagctcaaa tccttccact ctctgctccc
12540

ataggtctgg ggcttggcgt cccatgcttg cttccctgct aggtgcgaag ctcagggaag
12600
acgagtcagc atctaccttg ccgtctgccg tgttccctta ccatccccag cccagtgcag
12660
tagagtcagg gtctgtggct gacggcctga ttgccagacc ctgggcaagg tcctggggct
12720
tacagagagg aatcgggcac atccctgcca gcaactctta tggagcccag tggggcagct
12780
aaatcagcag agctgggatt tcccaatcct caggtcagca gcagagtcag gacctggggc
12840
tgggtgggca gcccccatga ctggctcagc taacagcgct gtgcccacca cagggagtca
12900
agtgattttt gaaggcccca ctctggggca gcaggaagac caagagcgga agaggttgtg
12960
taaggccatg accctgtgca tctgctacgc ggccagcatc gggggcaccg ccaccctgac
13020
cgggacggga cccaacgtgg tgctcctggg ccagatgaac gagtgagtcc ttggtcgcac
13080
cttctgggga caacgaagtg ggtaccgggg ctggagggac ctgcccacct ctctctgctc
13140
ctctgcagag tcctggaaag cctcggggca gccagacctg gcctgggagc ctggcagggg
13200
tggaaagatg tggccccatc tagcctctgt gtcctggcac ccctgtgccc acacagaagc
13260
cttagagagg atagggagct gatgtcaggg gagctaacgt cccagtctgc tttctgctat
13320
gatgcaagac ccaccacctc ccctggggtc agggactctg gctcagagag ggagtgtgga
13380
ttgaactctg agctaaagtc atggcagatg acaatgtact tccagacgct gggtccttgg
13440
ttgaaacttg tagaaaatag acacctctaa aagactcccc agcactccct ttgctcactg
13500
cttttggtgg ctaatggtga tggccccatg gcatccgagg tctacagatg gtatgaaggg
13560
ctggggttgg gtcattcact gcttcactgc ttcgttatag tccccttgtg aggtatcagg
13620
tgaaccatgg gatggtttgg aactttctag ccttggccac aaagggatgc aggccatgag
13680
gaccccaaga gggagagaaa cctgggccct gccgcggggt agtcatggtc tgttgagggt
13740
ggcaagatgc ctggggcttc caggcatgtc tggtacataa atgtactaat tgaggtatgt
13800
actaattgca gtgggcaggc acaaaaataa ggtgatgcca tcctttgcag acaggagcct
13860
ggacaggggt ggggagggca gtgggcgcag gagctgggag gtggaaagga caggtctgga
13920
gcctggctgg gcagaaacgt gaggttcaac aacccgtttg ttttaatttc gggagtgttt
13980
tctgtaatga tatccttaca gttctccagt aactttcttt gggaagagca gcccgtctgg
14040
gctgagtggg gaaagctctg cgcctgcttt gacactcttg agctaaaggg ggcgcccctg
14100
gggctagcag agccccgggg atgggaggcg gggcctgtgg tggaagtgac cctcctccag
14160
cctccgctct gggaagcttt tgagatttcc tttgctaagt ggggggaccg ttctttgcag
14220
aaacccacag agcgagattg ctgaggtctc tgcagatccc caaagatgtc agccaaatta
14280
catgcatgtg tataaaaggt gtatttttct ttttttttctt tttgagacaa gtctcgctct
14340
gtcgcccagg ctggagtgca gtggcgcgat gttggctcac tgcaacctct gcctcctggg
14400
ttcaagcgat tctcccgctt cagcctccct attagctggg attacaggcg cccgccacca
14460
tgcctggtta atttttgtat ttttagtgga gacggggttt caccatgttg gccaggccag
14520
tcttaagctc ctgaccttgt gccccacctg cctcggcctc ccagagtcct ggaattacag
14580

gcgtgagccc ctgcgcccgg ccacaaagtt gtatttttct ggagggatgg gccataactt
14640
ccatgagact cttagcaagg cctggacaca cagaagagtc agtgggtcat ttctcggcct
14700
tgtcttgtgc tgtggccatg ttctgaggct cccactcgat taggggacaa tgcttggcaa
14760
tggacttggt ggctagacct caggaggatg tggcctccac acaggcgcgc ctctcagggc
14820
ccagctgctg ctccgtcccc acgcacaggg ccaggctggc tcccacagct cagcatctga
14880
ggtgggggcc ggtgtcttct tgtaggttgt ttcctgacag caaggacctc gtgaactttg
14940
cttcctggtt tgcatttgcc tttcccaaca tgctggtgat gctgctgttc gcctggctgt
15000
ggctccagtt tgtttacatg agattcaagt aagtttgagc tgctcacagc ctaattatgc
15060
ctaattatgc ctcaaagctg cagaagagcc ctcagactca ataggcaggt ttacaaagtc
15120
cttcgtgtct ggccctgatc tttctccagc cctgtctcct gctagtctgc cctcctgttc
15180
cttcgaaccc aggctgctca ctgagctttg tgcacacgtg gtccccttc cctggaatgc
15240
cattctctac cttcccacct cctcagcctt caaggctagt tcaaatgctg cttccctgac
15300
ttttccccac ccccattcca tctctgagcg gcccctgggc atatcacagg cctgtccttt
15360
agtatctgca tttggcttcc ggtgactttg aattcctcca gaaccactct gatgctgggc
15420
accccgcaca gctcccagca cagggaggaa gagcaggcag gttaaagcaa ttaaagataa
15480
gctggtcccc acgtgccagt tcgacattgc tggacaagct tcctctttgc cgtgtgggtc
15540
catcaggcca ggtcaccgca aacctgtgac ttagctctga gctgagcgca tacgctctgt
15600
gcctcaatgc acggggagtt taagtcgagt aaaaccagca gtgattatga ccaaatccat
15660
ccaaacccag acatttactg aatacctctg gtgttcccag cagtgtacag gtcctagaaa
15720
gtttaccttc ctgttcctag cacacaggca agttcatcag gggtcacctt tgatggcagc
15780
cagactttgg acagaaacca tgacctgtgg ctgacaaata gctaaaaaaa agttattgtt
15840
tttctaaaac acacaaattt atctgtggtg caaaggtgat caggccacac caggatagaa
15900
agtactcagc tctgagttaa gtgcctgtgc tctgtgcctc catccacagg aagttcgagc
15960
caagtcaaac caggggaatt tgtgaccaga gggaagagac tgcagagctc agaggcaaaa
16020
gtgcccacgg aaacctgtga ttttgtgggg aaaataggga attttcctaa gttttcttct
16080
gaaggaggaa ctgttttgaa aactcccatt aaaaagttgc tatacaggcc gggcgcgatg
16140
gctcacacct gtaatcccaa catttttgga ggccgaggtg ggcagatcgc ctgaggtcag
16200
gagtttgtaa ccagcctggc caacatggtg aaaccccgtc tctactaaaa atacaaaaat
16260
tagccgggcg tggtagccca cgcctgaaat cccagcactt tgggaggcca aggagggcgg
16320
atcgcctgag gtcaggagct cgagaccagc ctggccaaca tggtgaaacc ccatctctac
16380
taaaaataca aaagttagct gggcatggtg gcacatgcct gtaaccccag ctacttggga
16440
ggctgaggca ggagaattgc ttgaagccgg gaggtagagg ttgcagtaag ccaagatcat
16500
gccactgcac tccagcctgg gcgacagagc aagactctgt ctcaaaacaa aaaaaaaagt
16560
tgctatacat attcaaaaca atcataataa tgatagtaag aatgacaata ttaatgatca
16620

ttgcccaaac cccactctgt cctgcccatg gacggggcag gggaaactgt ttgcatggct
16680
gcctggccac ccagcctggc tttgacagta gctctctttg ccctgcctct tgaatctgca
16740
ccagggccaa agtcctgttc atttgttcac atccgtcgaa caggtctctc aggagatggt
16800
cctgaacctg ctgcaggtga gcatctgtgt ctcctcatgg ggcaacagga ataataatga
16860
ccaacattta ttgagtgctc atcatgtgcc agacatgatt tcgagcgctc ttttcctttc
16920
tttattttat tttattttat tttatttatt tatttattta tttatttatt tatttattta
16980
tttatttttg agacagtgtc ttgctctgtc acccatgctg gagtgcagtg gtatgatctc
17040
ggctcgctgc aacctccacc acctgggttc aagcaattcc ccctgcctca gcctcccaag
17100
tagctggaat tacaggcacc caccaccacc atgcctggct aatttttgta ttttttagta
17160
gagatggggt tttgccatgt tggccaggct ggtcttgaac tcctaacctc cggtgatccg
17220
ccctccttgg cctcccaaag tgctggcgtt acagatgtga gccacctcgc ctggcccaag
17280
cactcttaaa cttaattaat tttcacaaca acctgcgagg tcagcactat tattattatt
17340
cccaatttac agacaaggca actgaggcat ggagaggtga tgtggtcaac acagagcttt
17400
gtaacaggga agtaggggga ctgagacttg aacccaggcc ctttggctcc cactgcatgg
17460
catcccctct tggggaggct gagggttgct gtccttagtt gcctccagac ctaagcatga
17520
ccaggtgtca gaaacactag ttggggccgg ggctgcccta gaaccccaag gcctactgag
17580
aaagaggagg gagatagcat ggcgccgagg ccgcaagggc accatcagct tcttgtctgg
17640
ccagaggcag atgtcaggcc cctggagact cacagccaga acctgaagct gagtccaccc
17700
agcctggcac ggccttcatc agcttttgtt gactggcggg ggagcctgag agtgtctgca
17760
gcagggggct tctgagcatg ctcgtggtgg ggtgcgtggc tgcagtccag tcccacccct
17820
tccccttccc gacgggccac tctagtttgg acgcatgcag tgtggctggc cggggtagct
17880
cacggcagct ttgttttggc tccagatctg gaaggtagag gacagctttt acattcggtt
17940
tgagtggtgg gaacagtgct ctggcccagg ccacgtcctg ccacaaacta agacctggtg
18000
gtccctgcct gcctttgtgg cctcatggac ctccccacct gaggccaggg agcacctgtc
18060
tcagcggcag gaggcagctc cactgtcagc tgttgctctc actagagttc ctcatctgaa
18120
cgatcctgga gaacgaggtt aagttcttgg cctctagcct aatccagaac aactatcttg
18180
ctgaagagcc tagtgcagcc tcctaggcta tatctagcca aaggggccag accccacccc
18240
aggaccacca agaactacat gggatattat tactggttat acctaactgt cccaaccagg
18300
cttacctcct gtaatagcca tgagggttct ttgggacccc tgccagggca gaggcatgca
18360
aagctcaaga atctctcccc tcttgttggc tctgcaacat attcagtcca agttcaccat
18420
ggtgcatcat ggtgaaggct gttctgctgc aggaggactc tgtggtcccc acccctgacc
18480
ctgacctagg cccctcacag gccaactgga tccatttact tgcatctcat gccagcctgg
18540
tcatcaccag atgaaattaa cccagagatg agagcaaagc tgctcagcac gagagactct
18600
gaaggcttgg cggtaccact gtggggcact ggcattggaa gactgcatac tccatgcagc
18660

cccagagtct gcagctactg tggtgttggg gatgagctgc cagcaccaaa tgcaggctct
18720
ggctcctggg ccactagtaa taccaaggtc acccctatg ctggaaacct gaagcccctg
18780
gctgagcccc agggtctcta ggacgacagt tggcagcaga gaggtgcttg gtagagcaca
18840
aactttacta agccaagggt gtggcagcag agaggccctg tcttacacca gcagagccat
18900
ccctgtgccg gatgtctaga gagtgtccct agcgggtgac cctcaggaca cacggccttg
18960
cccagcaggg agatcctagc cagccgtgta gacctgaggt cccatcagtg ttgcctcctt
19020
ttctgacccc tgagcacccc agaaagctgt gacctgatgt cctggtgtcc ccatgttcca
19080
ggccaagcca ccatcacacc aacacttggc cctcacactc tccaaggctg ttcacatcca
19140
gcactggctt caggaatgag ctcctattcc atcaacccct tccctcctat gattatgtct
19200
catggccccc gggaagggct ctcacgaggg agggctctcc aggacaatac tcttggcctt
19260
gcccacccct tcaaaccaac agtggctgga actggaatgt gtgaatggaa tattcagcat
19320
accttgaggc cttagtccta tgcacagtgg ccccagttat ccccctcca cagctgagct
19380
ccccttaca cctcctccaa gaacctcctc tcctccctgc ctcctcatgc caacgccacc
19440
ttaggggagg ccctgcagga caccctggac aatggacact ggtcccaagg gggcccatcc
19500
aggatggggg tgccatcctg ggctgtcttc cttcttgccc tagccatgct tgctgctaac
19560
cccagggtct cctggatccc taatcctgca cctccaactc cagggaacac aaggacccat
19620
tctgcccctg actagccctg tctgccaggg ttcatactca ctccctgcat ctccctgagc
19680
caccttggtg atgggggttg gcatcccaac accatcgaag gcagctccag gctgaggtgg
19740
aaggaggaag acttgggaag catgtgaggg agccctgttc ccaccttgcg caggctccga
19800
agctccttat ggccttcccc caggtgaccc tggagcagcc agtctccagg tgtctgggca
19860
cctgccgaga ccctctagcc tctctacaga gacttttttcc ctagtacatt ctgggatgga
19920
agaacaggag agggaaagag gcaggaaggg cctttctcca ggccccatag caggcgagga
19980
cagcattatg tgtctttttg ctacattctg ctgtagaaca tttaggctcc atctgaccag
20040
cacctgagcc aaccagtctg ccctgccctt ctctcatctt tgcattctcc agttttaaaa
20100
agtcctgggg ctgcgggcta gagagcaaga aaaacgagaa ggctgccctc aaggtgctgc
20160
aggaggagta ccggaagctg gggcccttgt ccttcgcgga gatcaacgtg ctgatctgct
20220
tcttcctgct ggtcatcctg tggttctccc gagaccccgg cttcatgccc ggctggctga
20280
ctgttgcctg ggtggagggt gagacaaagt aagtcttgga ttcaatagaa atcgctggct
20340
tagggccagg cgcgttggct cacacatgta atcccagcac tttgggaggc tgaggtgggt
20400
gggtcacttg aggtcaggag atcgagacca tcctggccaa catggtgaaa ccctgtctct
20460
actaaaaata cagaaaatta gcgaggcatg gtggcacatg cctgtagtcc cagctacttg
20520
ggagactgag gcaggagaat cacttgaacc caggaggcag aggttgcagt gagcccagat
20580
cgtgccactg cactccagcc tgggcaacag agagagactc cgtctcaaaa aaagagaaag
20640
aaagacacca ctggcttagt gcactagtgc ctaaatgctg ctggtctcgg ctacaggtgg
20700

caagaggaat gtgggccagg cactcatgct tggtcaagac ttttcctctt ttgggagctg
20760
ggtttcagag agcactctgt tggtttcatg actcattttt gttttctgac caagctccac
20820
aataagaccc taatgtgttc ctgtggtatc ctctcctccc tgagtaggct gagcagaaaa
20880
tccttggcca ggcagggtgg ccagagctgt gatgagagag atttcttggg ctaggagtag
20940
ggttcccaga gctctagttt ccaaatctct gctctgccat cttccctttc tcatcttcac
21000
atctggtcaa atccctccaa aggcacacat ctagggagct tcatagacag agacttggca
21060
aaggggggtac atgtagtttc tctcctggct aagacgttgt cagaatggaa gaaaggatga
21120
gaaacatgta catcctagaa aaggcagaag atgtgggcag ggagatgctg gtatgatggc
21180
catttcgttt tgaaggtcgg cttaggtcag caccaaagtc ttcatggtca ccctggtgaa
21240
cccagacaga attctagaga acctggtcaa gaagaggtcc tgaaatacac ttatggagaa
21300
tgcacgctga gaggggggaag taaactgctt aggatcaccc aaagttggtg gtcaagagtg
21360
tgggcatctt gatttctagc caggattcag tctcccatac cactcttatt tttttatttt
21420
tttgagacag agtctcacac tgtcacccag gctggagtgc aatggcatga tctcaactca
21480
ctgcaacctc cacctcccag caattctcct gcctcagcct gccgagtagc tgggattaca
21540
ggcgcccgcc agcatgtctg gctaattttt tgtatttttа gtagagacgg ggtttcacta
21600
tgttggccag gctggtcttg aactcctgac ctcgtgatcc gcccgcctca gcctcccaaa
21660
gtgctgggat tacaagtgtg agccactgca cctggccacc actcttgacc ttgacttta
21720
aggctgtgag cctgtttctt tgcatagaag catttggaca cagaactgcc ggagttgtga
21780
tgggtttgtt gagtgactgt ctctgtcgca gatgagctgt gcttttcccc acctaggtat
21840
gtctccgatg ccactgtggc catctttgtg gccaccctgc tattcattgt gccttcacag
21900
aagcccaagt ttaacttccg cagccagact gaggaaggta agtctcctgt tctgatcgcc
21960
cagtcatcag gactggagcc ctggaaccaa agggtcacta tgggatgcct tgggccctag
22020
agggagaaaa tcccatcata tccaagagga ttggctacaa aagcctggga aacagtggct
22080
ttcaagccac cggtggtatt atttagtgca aaatatcttt tttgcttttt aacatttgaa
22140
tttaacattt gaaattttat ttatattaca acaggaacag aaaatgtttc aaattttcca
22200
taacactgat ttccattcag cacaattttt tgttttctct cttcctccca gtctttgcta
22260
atatgcctgt atattacatt ataatcaaca cacacagttt gaatcctatt tgtttgttgt
22320
ttcttctacc acttttgatt gatattacac tataaacatt tcccactatt gctacagtct
22380
tcaaatatat tttctctaat aatggcatta tattgcgttg aggggttgta atcattctcc
22440
tgttatagaa cattttggct gttttgaatt ttttatttтc ataaattaat gttttcttgc
22500
atatagcttt tcctttgagg gtatttttтc tttaggataa acttctagga gtaatattgc
22560
tggggtgata gaatacaaag tcttaatggc ccttaaaatg tatggccaaa ttgcttttca
22620
aaaaggtcat accaatttac gatgctattg gcagtgtgtg taatagtttg atcatatcct
22680
caccagcaat gtatatatta ttgtaaactt tagctaattt ataagtagga gatggtacct
22740

cattgtcctt attagcttta ttccccttg attagatttc ttttgtcttc taatgctgct
22800
ctggtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtttctt tttcttccca
22860
gaaaggaaaa ctccatttta tccccctccc ctgctggatt ggaaggtaac ccaggagaaa
22920
gtgccctggg gcatcgtgct gctactaggg ggcggatttg ctctggctaa aggatccgag
22980
gtaacttctc cagccacagg ctgcccagag ccctcttctt cgtcaagagg gtggcgtttc
23040
tccacccttc catccctggg cttgtgtgtt tctgtgcctg catccttcgt ataaccgcac
23100
attccttgag gacatggact ctgtcttgtc atctaggaac tctacaccac acacagggcc
23160
tggaagacag aaagtaacct ttgagcgatt gcaggaatga gtgaatgagt gaccgtggtt
23220
agccaagaga ggcagaggac actgtcagtt accctctggg gcttgatcac aataatctct
23280
gctttgattt gtctgaggga aatctttctt tccaatcctt gtcaatattg tttgctacta
23340
cttttggtcc ttctactggc tacttaacat ggtagctact tcaaaatttt tctttagcta
23400
agtatgtagc agcgtaggag gtgaggaaca tgttggaaaa cacacaaaaa tataactttc
23460
tttacctcct tctttccctc ctggggaaga aatgagccag agggagggat gagctagctt
23520
gctgctgctg tcctccaacc aaccatctac ctacccaagt atccaggagt gtaatagaca
23580
gacttggtct agttattgct gtttcttcaa tatctaggac acagcctggt gcctagtggg
23640
tgctaagttt ttgcggaggt gaacaaatcc atccatctag tcacctctcc atccatcatc
23700
catgcatcta ttcatgtctt catccatcca tccgtccgtc catccgtcca tccatccatc
23760
catccatcca tccatccatc catccatcca tctgtccatc catccaccca tctatctatc
23820
cacccatcta tctatccatc catccatcca cccatctatc catccatcca tccatccatc
23880
catccatcca tccatccatc tatccatcca tccatccaac attcctatta tgtccctagt
23940
gttatgccag gcacagagat tacagaggag attgagatac ggtccctgtt cgtggcagac
24000
ttcacagact agggaggggc acatatatga aagggcattt caggaagtag cacacgagca
24060
agggaaaaat gtgaggtatt tagctgagga gaagtagaag atgaggctgg taaggctacc
24120
agaagccact tctctgaggg ccccaagata gaggggtgtg gacttgatcg tgaatgcagt
24180
agacagccac tgaaggactg aggccagggg gtgagttggt cagatctgca catgaggaaa
24240
tcactctgat gtctggagtg ggggcctggg ctgggcaggg cttggaggag aactagctga
24300
gactctgcag ccttccatct cactcaggct cagaactttg gactctgtgg acattctctc
24360
ctcctttggc ccccagctca gcacagtctc cagctttact tcggactcag actattcctg
24420
ctcagccttc gttgctgact tctctgttct ccctgaaaca ggagtgtctg cccaggctct
24480
gtccttggcc tctcctcttt ttacacttca ttctctccct ggacaatctc ttctcagccc
24540
aaagccctaa atctaaacct tcaatttctg gttgaaatca ttctcctgag cttccaaaac
24600
tgtggagcac tgaagaggag gagatggatg tgagacattt gggtgacttg gtgactgact
24660
gggtataagg aaggagggga acagagaccg gcagcatgac tcccagcctg ctgggctgga
24720
tggctggtgg atggtgagtc cattcaccaa actgggaggc ccagagagag aagcagattc
24780

tgggctatgg aggatgaatg cagggtggag catgttgagt ctgttgtgct cttgggacat
24840
ctggatggac atttccagaa ggcatatggg tatgtaaatc cacatagtag gccagctggc
24900
tggaaataca gatttaggag acagcagagt gaggacgggg atgaaaatgg tgggaatgga
24960
tgaggtcacc tatgaggtgt agagagagag ggtcgggagg ggaggatggg ccaagcttgc
25020
cctggccctg aaggaactgc aagctgggag cgctgagatg actgccctcc tggtgcttcc
25080
caggcctcgg ggctgtccgt gtggatgggg aagcagatgg agcccttgca cgcagtgccc
25140
ccggcagcca tcaccttgat cttgtccttg ctcgttgccg tgttcactga gtgcacaagc
25200
aacgtggcca ccaccacctt gttcctgccc atctttgcct ccatggtaag taacctgaca
25260
gtggggagga gcccttccat ttcacaggaa cacatggcca tattgtgggt ccctgacgag
25320
gcagcaatgt ccaggccaga ctcagaccag gctttggaga cccaggtctg actgtgacgt
25380
ggatttgtgg accctggatg cctctgcccc tgaggcctcc actgctttgc cactcctctt
25440
tgtacccctc ctgctgacca aagcaccaac catggaccaa gtgctcaaat ttattttata
25500
aatctaattg gattattttt caagctgggg agacaggact tgggctaagg aggagcaggc
25560
cagtgccgtg gtctctgagc atgtagcaca ggtgtgcagg aggactgcag actgggagca
25620
ccactggctg gaaaccccag gaagaggcct tggaggagtg gggacttggg agtaggtagg
25680
aagggagaga gaattctggg aagatggagc agcacaagga aaggcaatgg tgcacatgac
25740
tgaggactcc tggaagcctg gcttggtgag cacagggata agggatcctg gggagtggag
25800
agaggtagct gtcggttgtg ggaaaagctg ctgagtgcca ggctaaggca ttctgttcta
25860
tggactagca tgtttttag ttgggagtta gaagaaagca gagcttatag gaaaatcagt
25920
ggctatggtt ttttttttt ttttttttt ttttttttt atgcatttcc ttctgtcatc
25980
cattgcaaag acgtaccagc ttcagggtag tatggaaaga tccctggtct cgcagtcaga
26040
agacccgagt tcaagatgtg ggatctctga acatggccct tcagttcttt cttccgagag
26100
ctgtgctgat ggccaagtaa gatgagggct atgaaaagcc tctgtagact gcaaaatgag
26160
catgggagag gctgtcatta ttctggaatt gggagacaga tttacagagg gcctgaacac
26220
aggattgaag gtggtgaatt tccattcggc tgcctgggcg tctgcatgta taaaaagcaa
26280
acctaagtgg ttttttttctc ctccaagtga agatgaaagt gttaaaaata gcaaggaggt
26340
gaaagtgttc aaaatagcaa agtggcctgt ctcctcttct cctaagcaga ctgtccaaac
26400
agacgcccag tagaaggagc acctttgat actgggcacg tggtggtgat gcctcctctc
26460
tcctagcaca ggccgtggct ttgtcatctc cagccctaac tgggagcacc gagggttcca
26520
accaggcaaa tgcaggccct aacgggctct ttgaaaacgg gcttttctag aaccaggaac
26580
ctcaagtaaa aactccccca gctacctcta aggcccatca cactcctgtc tcacgcccac
26640
ctatgagaaa ggaaaaggtg atggtcattg agctgggctg cagaggagtg tgaggtgcag
26700
acaccatgag gtacccacag ccaggaaaac gaggatggtc ggggagacgc gccagcgaag
26760
agctgggccc ctgcgtggga cccctcagtg gttcccaggg ggcgtgggac ttgcgcagtc
26820

ctttcagagg gctgtttacc aacaggaacc gtaacattaa acctgctcag acccccttgac
26880
tcagcaattt catgtctggg aatatatctt aggaaaataa tcagagatgc ctaccaacat
26940
atgtgatgat gatgtatgac agaattatta tacaaatata tccatagtaa caggggggttt
27000
gctgaaataa attatcatat attcatataa tatgacatta tcaggccatt aaaaatcaca
27060
gtttcaaaga gtaataaaat gggaacatgc tcatagtata gtttttttaaa attgcagatg
27120
gtatatggct aaaaatgtct aataatgcaa agatgtatac agaccttaat cctctagcct
27180
cctccctaga gatgacctct gttaatttct caaatatttt tctggatatt ttacacactc
27240
acacactttt tttgagacag agtttcactc ttgtcaccca ggctggagtg caatggtgtg
27300
atcttggctc actgcaacct ccacctcccg ggttcaagag attctcctgc ctcagcctcc
27360
cgagtagctg ggattacagg tgcctgccac cttgcctggc taatttttttg tatttttagt
27420
agagacgggg tttcaccaca ttggtcaggc tggtctcaaa ctcctgacct caggtgatcc
27480
gcctgccttg gcctcccaaa gtgctgggat tacaggcgtg agccactgcg cccggccatt
27540
catcttaatt tttaaaaaat ctaaccatga agccttggtt atcttggaga gctttcctga
27600
ttagcacaaa aagaaaaaaa aatccaattc tttacagctg catactattc cattatttgt
27660
atgtgtcata ttttatttaa ccatcctgct attagtgacc attgagttgg cttcctgtgt
27720
tttgccgtta catggttgca acaaacatgt ttgcatgtgt ctgccctcat gtgcatgata
27780
catgattgat ttgatagatt ttaggaatta catcattcat tcatacactc agcaaatatt
27840
taatgagtgc ctactctctg ataggtgctg ttggatgtgg ctaaatttta aagtgtagaa
27900
tttaaaaggt ggctaccaaa ttccatgtgc aaaatgaccc cacgcatgta taaaaacaca
27960
cacatccaca gatttatatg cgggagagaa gatgtggtcc ctggcctcta ggctctctca
28020
gtctgtggca agacagacag acatgtgcac gcggcactgt aaggttgagc acagtctaag
28080
tactcagcat ggtctctggc acatagtagg tgcccaagaa atacatgtcg aatgaattga
28140
gggggtaagg ccttctaggg caggtggcct ctgacctcag ccttcagtgt tccgtaggtg
28200
gaattatctg ccagagacgt ggcaaaaggg agaggaacca agactgaggc acagaggttc
28260
aaacgtaccc ggcacattca gagaatcctt ttcagaatca cgtccccaag agcttctgtg
28320
ttctgtacgg tgatgttgca gtgctgtttt tccgcagtct cgctccatcg gcctcaatcc
28380
gctgtacatc atgctgccct gtaccctgag tgcctccttt gccttcatgt tgcctgtggc
28440
caccctcca aatgccatcg tgttcaccta tgggcacctc aaggttgctg acatggtaac
28500
acagctgttt ttatttactc ccgtcggact ataacgctgt tgtcataagg gatgccccat
28560
ttatgaatga cagagtttca aaacgatgtc atgtgacttg ggaatgccac ggaacatcca
28620
gacctgtagc cattgttgac atttataatg cagcttttct tctttttctg agatgatctc
28680
aagcctcaca cactgttctt tctctgaggt gggttataga ctctcccacc tggagaagcc
28740
tgtgcaggca ccaggggagt ccttggaagg ggtgaaggtg gggctgaggg actcatatgg
28800
ccaaggatga acttgacaaa ttagcaagaa ccatgaagat aggcagggca ggcttaggca
28860

gcaggggat gctaatgaca gtcacagaga tttgtagggg tgcctgaaga ggtagaagca
28920
gggagaggga gagagagagc actgcctggg agtagatgat gccttggaaa caaatgtagt
28980
cagaggaaga actcttcatt agctctgtca cctttgctgg gagaagggca gctttgcagc
29040
tctgggctgg gaaagaggca agtgtttgag cccaagaggc cagaaatgta cctgggacca
29100
atcgggtgtt cgttatctca gagcctctgc tgggtatctc agggactcca tgagcatttt
29160
caaaaaaaaa ggtgggtccc agaaaccatg gactgcaaac ttgactccaa tcccccagta
29220
aaatatctac aacagggtag tgaagcgatg gttagtgacc atgagggaag cttgcagagc
29280
aggcatcaga aagagcctga ggaggtccac agggaagctg gcacgtcctt gtaggatagt
29340
taaggcactg gggtgagcaa tgaacctgga ctcacggaac actgggctct gtgaccgttt
29400
ccctgaatgg cctaagctgt tgcctcctgt cacttctctg aggtcatttt ccaaatgcgc
29460
acgggcatag agaacccatc cactctgcct acttcccagg gatgccttga gcactgagga
29520
tacctggggg acatgaagtc gcactgtcct gggggtcggg acaccccagc cagggacaga
29580
gcatggcaca gggacatcga ggcccagtga gccgaccctt tgtcctcctc tctgagagca
29640
ctagtcccca gcaggcctca gggtgctgac tctgtctctt ttccaggtga aaacaggagt
29700
cataatgaac ataattggag tcttctgtgt gtttttggct gtcaacacct ggggacgggc
29760
catatttgac ttggatcatt tccctgactg ggctaatgtg acacatattg agacttagga
29820
agagccacaa gaccacacac acagcccttа ccctcctcag gactaccgaa ccttctggca
29880
caccttgtac agagttttgg ggttcacacc ccaaaatgac ccaacgatgt ccacacacca
29940
ccaaaaccca gccaatgggc cacctcttcc tccaagccca gatgcagaga tggtcatggg
30000
cagctggagg gtaggctcag aaatgaaggg aacccctcag tgggctgctg gacccatctt
30060
tcccaagcct tgccattatc tctgtgaggg aggccaggta gccgagggat caggatgcag
30120
gctgctgtac ccgctctgcc tcaagcatcc cccacacagg gctctggttt tcactcgctt
30180
cgtcctagat agtttaaatg ggaatcggat cccctggttg agagctaaga caaccaccta
30240
ccagtgccca tgtcccttcc agctcacctt gagcagcctc agatcatctc tgtcactctg
30300
gaagggacac cccagccagg gacggaatgc ctggtcttga gcaacctccc actgctggag
30360
tgcgagtggg aatcagagcc tcctgaagcc tctgggaact cctcctgtgg ccaccaccaa
30420
aggatgagga atctgagttg ccaacttcag gacgacacct ggcttgccac ccacagtgca
30480
ccacaggcca acctacgccc ttcatcactt ggttctgttt taatcgactg gccccctgtc
30540
ccacctctcc agtgagcctc cttcaactcc ttggtcccct gttgtctggg tcaacatttg
30600
ccgagacgcc ttggctggca ccctctgggg tccccctttt ctcccaggca ggtcatcttt
30660
tctgggagat gcttcccctg ccatccccaa atagctagga tcacactcca agtatgggca
30720
gtgatggcgc tctgggggcc acagtgggct atctaggtcc tccctcacct gaggcccaga
30780
gtggacacag ctgttaattt ccactggcta tgccacttca gagtctttca tgccagcgtt
30840
tgagctcctc tgggtaaaat cttccctttg ttgactggcc ttcacagcca tggctggtga
30900

caacagagga tcgttgagat tgagcagcgc ttggtgatct ctcagcaaac aacccctgcc
30960
cgtgggccaa tctacttgaa gttactcgga caaagacccc aaagtggggc aacaactcca
31020
gagaggctgt gggaatcttc agaagccccc ctgtaagaga cagacatgag agacaagcat
31080
cttctttccc ccgcaagtcc attttatttc cttcttgtgc tgctctggaa gagaggcagt
31140
agcaaagaga tgagctcctg gatggcattt tccagggcag gagaaagtat gagagcctca
31200
ggaaacccca tcaaggaccg agtatgtgtc tggttccttg ggtgggacga ttcctgacca
31260
cactgtccag ctcttgctct cattaaatgc tctgtctccc gcggaaagct ccactgtgct
31320
gctgacttgt ctctggtttt ctgcagtgtg gggagcccag ggaggtggat gaatgaacag
31380
ttagttacgc cctgcccacc tgctgggtgc caggccttcc tgtccctgtt gaatccacta
31440
gttatctcgt tcgatctttg cagcaaccct gtgagatagg aaggtgttat tatcttgctt
31500
tgtctttcaa aaaaaaagcg aggctcaggg aggccaaggg aagtgtccaa agtcacacat
31560
caagttactg gcagttacag ttccaaccaa gagcttccaa ctccataccc cctgctcctt
31620
ctgctagcca tgaagggctt tggccttata gggcttgtag ggaaaggtga gtggccaaga
31680
gcaagtccat gccaagggaa gatctccaaa catgagtccc tgtctgttgc ctcccctgag
31740
ataggcacag gacaagtgat caatgagaca gggtggtcct tgccctaaga agcaaagtgt
31800
ttggttgggg agggaagtag ggaaaaggct gccacctccc cccaccaagg tacaactgct
31860
gacttccttc ctccccagcc ctctatcact gccctctgtg ccgctgccgt tgactggcct
31920
gccccaccag actgagggct ctgactgccc accgagtcta gtgtcagcat tatggctgac
31980
ccagagcagg ctatacagtt agtatgatgg ataaataaat gattggtcag tgcagtcaat
32040
taggtgcaag ctgttggtag taggcaaggt caatgaaggt catccaaggt gggcattgaa
32100
ggatgagtag aatggccagg ggtaatgggg gaggaactgg tgggtgggtg gaggactctt
32160
ccagacacca tgtggttgag ggctgacaaa aagctgggtg gagggcttcc agagtgccaa
32220
gctcccacct gaagaggctg accagaggcc aatcctaaac aactctaggt gttggctgga
32280
gttgcactaa agtgtatggc ctccccaacc aaaccctttg cttcttaggg caaggaccac
32340
cctgtctcat tgatcactgt cctgagccta tctcagggga ggcaaaagag agggacctgt
32400
attcagagat cttcccttgg catgacttgc ttttggccac ttacctttcc ctacaagctc
32460
tatgaggcca aggcccttca tggttagtgt aaggagcagt gggcatggag ttggaagatc
32520
tgggttggaa cggtaactgc cactaactcg atgtgtgatt ctgaacactt aacttagcca
32580
tacatgctct cttatttgct tttgatggca aataagagaa ggcccagcaa acagtggctt
32640
aaaccagaag gtcaattaat gtttactttt caggaagtct gtaggtagat ggttgctggc
32700
attggcccaa cagctcattt cagcctccaa ggacttgcgc tccatagtcc actctgtcat
32760
cttaaagcct tcacactttt acccccatgc ttgaccccca ggctacatac acagct
32816

<210> 4
<211> 619
<212> PRT

<213> Xenopus laevis

<400> 4

```
Met Val Ser Ile Gly Lys Trp Ile Leu Ala Asn Arg Asn Tyr Phe Ile
 1               5                  10                  15
Ile Phe Leu Val Pro Leu Phe Leu Leu Pro Leu Pro Leu Val Val Pro
            20                  25                  30
Thr Lys Glu Ala Ser Cys Gly Phe Val Ile Ile Val Met Ala Leu Phe
        35                  40                  45
Trp Cys Thr Glu Ala Leu Pro Leu Ala Val Thr Ala Leu Phe Pro Val
    50                  55                  60
Leu Leu Phe Pro Met Met Gly Ile Met Asp Ser Thr Ala Val Cys Ser
65                  70                  75                  80
Gln Tyr Leu Lys Asp Thr Asn Met Leu Phe Ile Gly Gly Leu Leu Val
                85                  90                  95
Ala Ile Ser Val Glu Lys Trp Asn Leu His Lys Arg Ile Ala Leu Arg
            100                 105                 110
Val Leu Leu Ile Val Gly Val Lys Pro Ala Leu Leu Leu Leu Gly Phe
        115                 120                 125
Met Val Val Thr Ala Phe Leu Ser Met Trp Ile Ser Asn Thr Ala Thr
    130                 135                 140
Thr Ala Met Met Ile Pro Ile Ala Gln Ala Val Met Glu Gln Leu His
145                 150                 155                 160
Ser Ser Glu Gly Lys Val Asp Glu Arg Val Glu Gly Asn Ser Asn Thr
                165                 170                 175
Gln Lys Asn Val Asn Gly Met Glu Asn Asp Met Tyr Glu Ser Val Met
            180                 185                 190
Pro Ser Gly Lys Met Ala Leu Ala Ile Asp Asn Thr Tyr Ala Thr Glu
        195                 200                 205
Asn Glu Gly Phe Glu Ile Gln Glu Lys Ser Thr Lys Asp Pro Glu Pro
    210                 215                 220
Ser Lys Gln Glu Lys Gln Ser Ile Gly Pro Ile Val Ile Glu Pro Glu
225                 230                 235                 240
Asp Glu Lys Gln Thr Glu Glu Lys Gln Lys Glu Lys His Leu Lys Ile
                245                 250                 255
Cys Lys Gly Met Ser Leu Cys Val Cys Tyr Ser Ala Ser Ile Gly Gly
            260                 265                 270
Ile Ala Thr Leu Thr Gly Thr Thr Pro Asn Leu Val Met Lys Gly Gln
        275                 280                 285
Met Asp Glu Leu Phe Pro Glu Asn Asn Asn Ile Ile Asn Phe Ala Ser
    290                 295                 300
Trp Phe Gly Phe Ala Phe Pro Thr Met Leu Val Leu Leu Ala Leu Ser
305                 310                 315                 320
Trp Leu Trp Leu Gln Phe Ile Tyr Leu Gly Val Asn Phe Lys Lys Asn
                325                 330                 335
Phe Gly Cys Gly Gly Asn Ala Glu Gln Lys Glu Lys Glu Lys Arg Ala
            340                 345                 350
Phe Arg Val Ile Ser Gly Glu His Lys Lys Leu Gly Ser Met Thr Phe
        355                 360                 365
Ala Glu Ile Ser Val Leu Val Leu Phe Ile Leu Leu Val Leu Leu Trp
    370                 375                 380
Phe Thr Arg Glu Pro Gly Phe Met Pro Gly Trp Ala Thr Ile Ser Phe
385                 390                 395                 400
Asn Lys Gly Gly Lys Glu Met Val Thr Asp Ala Thr Val Ala Ile Phe
                405                 410                 415
Val Ser Leu Met Met Phe Phe Phe Pro Ser Glu Leu Pro Ser Phe Lys
            420                 425                 430
Tyr Gln Asp Thr Asp Lys Pro Gly Met Lys Pro Lys Leu Arg Val Pro
        435                 440                 445
Pro Ala Leu Leu Asp Trp Lys Thr Val Asn Glu Lys Met Pro Trp Asn
    450                 455                 460
Ile Val Ile Leu Leu Gly Gly Gly Phe Ala Leu Ala Lys Gly Ser Glu
465                 470                 475                 480
Glu Ser Gly Leu Ser Leu Trp Leu Gly Glu Lys Leu Thr Pro Leu Gln
                485                 490                 495
Ser Ile Pro Pro Ala Ala Ile Ala Leu Ile Leu Cys Leu Leu Val Ala
            500                 505                 510
Thr Phe Thr Glu Cys Thr Ser Asn Val Ala Thr Thr Thr Leu Phe Leu
```

```
        515                 520                 525
Pro Ile Leu Ala Ser Met Ala Lys Ala Ile Gln Leu Asn Pro Leu Tyr
    530                 535                 540
Ile Met Leu Pro Cys Thr Leu Ser Ala Ser Leu Ala Phe Met Leu Pro
545                 550                 555                 560
Val Ala Thr Pro Pro Asn Ala Ile Ala Phe Ser Tyr Gly Gln Leu Lys
                565                 570                 575
Val Ile Asp Met Ala Lys Ala Gly Leu Leu Leu Asn Ile Leu Gly Val
            580                 585                 590
Leu Thr Ile Thr Leu Ala Ile Asn Ser Trp Gly Phe Tyr Met Phe Asn
        595                 600                 605
Leu Gly Thr Phe Pro Ser Trp Ala Asn Ala Thr
    610                 615
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/046407 A3

(51) International Patent Classification[7]: C12N 15/12, 15/11, 15/63, 1/21, 5/10, C07K 14/705, 16/28, A01K 67/027, C12Q 1/68, G01N 33/68

(21) International Application Number: PCT/US01/45661

(22) International Filing Date: 5 December 2001 (05.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/729,094 5 December 2000 (05.12.2000) US

(71) Applicant: **PE CORPORATION (NY)** [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: **CHATURVEDI, Kabir**; c/o Celera Genomics Corporation, 45 West Gude Drive C2-4 #21, Rockville, MD 20850 (US). **WEI, Ming-Hui**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **KETCHUM, Karen, A.**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **DIFRANCESCO, Valentina**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **BEASLEY, Ellen, M.**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
13 March 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/046407 A3

(54) Title: ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the transporter peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the transporter peptides, and methods of identifying modulators of the transporter peptides.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/45661

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C12N15/11 C12N15/63 C12N1/21 C12N5/10
C07K14/705 C07K16/28 A01K67/027 C12Q1/68 G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A01K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, GENSEQ, SEQUENCE SEARCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
L	US 2002/019028 A1 (BEASLEY ELLEN M ET AL) 14 February 2002 (2002-02-14) L:priority figures 1,2	1-23

L	DATABASE GENSEQ 'Online! 11 January 2002 (2002-01-11) TANG YT ET AL: "Human NaDC-2 homologue-encoding cDNA, SEQ ID NO:1284" Database accession no. ABA09508 XP002222098 L document cited to provide information on the relevant sequence disclosed in WO 01 57188 the whole document	1-5,8-16
P,X	& WO 01 57188 A (HYSEQ INC ;LIU CHENGHUA (US); TANG Y TOM (US); DRMANAC RADOJE T (US)) 9 August 2001 (2001-08-09)	

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 November 2002	11/12/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Espen, J

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
L	DATABASE GENSEQ 'Online! 11 January 2002 (2002-01-11) TANG YT ET AL: "Human NaDC-2 homologue, SEQ ID NO:2634" Database accession no. ABB12264 XP002222099 L document cited to provide information on the relevant sequence disclosed in WO 01 57188 A the whole document	1-5,8-16
P,X	& WO 01 57188 A (HYSEQ INC ;LIU CHENGHUA (US); TANG Y TOM (US); DRMANAC RADOJE T (US)) 9 August 2001 (2001-08-09)	

E	WO 02 12340 A (INCYTE GENOMICS) 14 February 2002 (2002-02-14) SEQ ID NOs 25,55	1-5, 7-16, 19-23

P,X	WO 01 57274 A (CHEN WENSHENG ;HANZEL DAVID K (US); PENN SHARRON G (US); RANK DAVI) 9 August 2001 (2001-08-09) SEQ ID NOs 15609,25711, and 26073	1-6,8-16

P,X	DATABASE GENSEQ 'Online! 16 October 2001 (2001-10-16) TANG YT ET AL: "Human protein encoding cDNA sequence SEQ ID NO:448" Database accession no. AAH99613 XP002222100 PX document cited to provide information on the relevant sequence disclosed in WO 01 53455 the whole document	1-5,8-16
P,X	& WO 01 53455 A (HYSEQ INC ; LIU CHENGHUA (US); TANG Y TOM (US); DRMANAC RADOJE T (US)') 26 July 2001 (2001-07-26)	

P,X	DATABASE GENSEQ 'Online! 16 October 2001 (2001-10-16) TANG YT ET AL: "Human protein sequence SEQ ID NO:1187" Database accession no. AAM25672 XP002222101 PX document cited to provide information on the relevant sequence disclosed in WO 01 53455 A2 the whole document	1-5,8-16
P,X	& WO 01 53455 A (HYSQ INC; LIU CHENGHUA (US); TANG Y TOM (US); DRMANAC RODOJE T (US)) 26 July 2001 (2001-07-26)	

	-/--	

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	SEKINE T ET AL: "Expression cloning and characterization of a novel multispecific organic anion transporter" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 272, no. 30, 25 July 1997 (1997-07-25), pages 18526-18529, XP002097953 ISSN: 0021-9258	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/45661

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 17,18
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 17,18

Present claims 17 and 18 refer to an agent defined by reference to a desirable characteristic or property, namely that binds to any of the peptides of claim 2.

The claims cover all agents having this characteristic or property, whereas the application provides no support within the meaning of Article 6 PCT and no disclosure within the meaning of Article 5 PCT. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the agent as a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search impossible.
Consequently, no search has been carried out for claims 17 and 18.

Nucleic acid molecules relating to SEQ ID NO 3 having a length of 32816 nucleotides, were not searched,
since it is not apparent what is the relationship between the human transporter proteins claimed and this lengthy nucleic acid sequence.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/45661

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2002019028 A1	14-02-2002	AU 3944402 A	18-06-2002
		WO 0246407 A2	13-06-2002
WO 0157188 A	09-08-2001	AU 2591801 A	31-07-2001
		AU 3128801 A	14-08-2001
		AU 3297101 A	07-08-2001
		AU 3300301 A	07-08-2001
		AU 3329301 A	14-08-2001
		AU 3484701 A	14-08-2001
		AU 3484801 A	14-08-2001
		AU 3486501 A	14-08-2001
		AU 3494401 A	14-08-2001
		AU 3665801 A	14-08-2001
		AU 3666001 A	14-08-2001
		AU 3666301 A	14-08-2001
		AU 3672101 A	14-08-2001
		AU 4314201 A	14-08-2001
		EP 1254265 A1	06-11-2002
		EP 1254266 A1	06-11-2002
		EP 1254268 A1	06-11-2002
		EP 1254269 A1	06-11-2002
		EP 1254270 A1	06-11-2002
		WO 0153326 A1	26-07-2001
		WO 0155334 A2	02-08-2001
		WO 0155335 A2	02-08-2001
		WO 0157255 A1	09-08-2001
		WO 0157260 A1	09-08-2001
		WO 0157175 A2	09-08-2001
		WO 0157261 A1	09-08-2001
		WO 0157262 A1	09-08-2001
		WO 0157187 A2	09-08-2001
		WO 0157265 A1	09-08-2001
		WO 0157188 A2	09-08-2001
		WO 0157266 A1	09-08-2001
		WO 0157267 A1	09-08-2001
		WO 0157190 A2	09-08-2001
		US 2002128187 A1	12-09-2002
		US 2002137044 A1	26-09-2002
		US 2002127199 A1	12-09-2002
WO 0212340 A		NONE	
WO 0157274 A	09-08-2001	AU 3087801 A	14-08-2001
		AU 3087901 A	14-08-2001
		AU 3088001 A	14-08-2001
		AU 3088101 A	14-08-2001
		AU 3088201 A	14-08-2001
		AU 3088301 A	14-08-2001
		AU 3275701 A	14-08-2001
		AU 3275801 A	20-11-2001
		AU 3275901 A	14-08-2001
		AU 3276001 A	14-08-2001
		AU 3311401 A	14-08-2001
		AU 3658901 A	14-08-2001
		AU 6343201 A	11-12-2001
		AU 9295701 A	02-04-2002
		GB 2361238 A ,B	17-10-2001
		GB 2373500 A	25-09-2002

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/45661

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0157274 A		GB 2374929 A	30-10-2002
		GB 2375539 A	20-11-2002
		GB 2375111 A	06-11-2002
		GB 2374872 A	30-10-2002
		WO 0157270 A2	09-08-2001
		WO 0157271 A2	09-08-2001
		WO 0157272 A2	09-08-2001
		WO 0157273 A2	09-08-2001
		WO 0186003 A2	15-11-2001
		WO 0157274 A2	09-08-2001
		WO 0157275 A2	09-08-2001
		WO 0157276 A2	09-08-2001
		WO 0157277 A2	09-08-2001
		WO 0157278 A2	09-08-2001
		WO 0157251 A2	09-08-2001
		WO 0157252 A2	09-08-2001
		WO 0192524 A2	06-12-2001
		US 2002081590 A1	27-06-2002
		US 2002048763 A1	25-04-2002
		US 2002048800 A1	25-04-2002
		US 2002123474 A1	05-09-2002
		GB 2360284 A ,B	19-09-2001
		EP 1158049 A1	28-11-2001
		US 2002102252 A1	01-08-2002
		AU 9481201 A	08-04-2002
		WO 0226818 A2	04-04-2002
WO 0153455 A	26-07-2001	AU 2292401 A	31-07-2001
		AU 2591801 A	31-07-2001
		AU 2593601 A	31-07-2001
		AU 2595501 A	31-07-2001
		AU 2596501 A	31-07-2001
		AU 2598301 A	31-07-2001
		AU 2728401 A	31-07-2001
		AU 2734401 A	31-07-2001
		AU 2734801 A	31-07-2001
		AU 2738501 A	31-07-2001
		AU 3265701 A	31-07-2001
		EP 1242596 A1	25-09-2002
		EP 1240178 A2	18-09-2002
		EP 1242580 A1	25-09-2002
		EP 1242443 A1	25-09-2002
		EP 1250346 A2	23-10-2002
		EP 1254256 A2	06-11-2002
		EP 1248848 A1	16-10-2002
		WO 0153312 A1	26-07-2001
		WO 0153453 A2	26-07-2001
		WO 0153326 A1	26-07-2001
		WO 0153454 A2	26-07-2001
		WO 0153455 A2	26-07-2001
		WO 0153456 A2	26-07-2001
		WO 0153466 A1	26-07-2001
		WO 0152616 A2	26-07-2001
		WO 0153500 A1	26-07-2001
		WO 0153515 A1	26-07-2001
		WO 0153485 A1	26-07-2001
		US 6465620 B1	15-10-2002
		US 2002146692 A1	10-10-2002